

# Multiscale modelling of bionano interface

Hender Lopez<sup>1</sup>, Erik Brandt<sup>2</sup>, Alexander Mirzoev<sup>2</sup>, Dmitry Zhurkin<sup>2</sup>, Alexander Lyubartsev<sup>2</sup>, and Vladimir Lobaskin<sup>1</sup>

<sup>1</sup>*School of Physics, Complex and Adaptive Systems Lab,  
University College Dublin, Belfield, Dublin 4, Ireland*

<sup>2</sup>*Department of Materials and Environmental Chemistry,  
Stockholm University, SE-10691, Stockholm, Sweden*

## I. INTRODUCTION

Over the last decade, in vitro and in vivo experiments have produced significant amount of veritable information that can be integrated into theoretical models with the aim of predicting possible health and environmental effects of engineered nanoparticles (NP) [1]. However, even the most systematic studies leave the question of precise toxicity mechanisms associated with nanoparticles wide open [2–4]. An important finding arising from these studies is that the toxic effects can emerge either from membrane damage or from interaction of nanoparticles, once they are inside the cell, with the internal cell machinery. Therefore, an evaluation of possible risks should include an assessment of nanoparticle ability to penetrate, modify, or destroy the cell membrane [4]. Being selectively permeable, membranes participate in control of the transport of vital substances into and out of cells. Whereas some biomolecules may penetrate or fuse with cell membranes without overt membrane disruption, no synthetic material of comparable size has shown this property [5]. Among the factors determining the outcome of NP-membrane interaction the surface properties of nanomaterials play a critical role, which can implicate the membrane or plasma proteins in conditioning NP prior to cell penetration.

The detailed understanding of the crucial stages of NP-cell membrane interaction can be achieved with computer simulation. Molecular dynamics is now a well-recognized tool for studying intermolecular interactions, self-assembly, and structure of biomolecules or their complexes. The reliability and predictive character of molecular modelling has improved significantly during the last few years, with development of new, carefully parameterized force fields, simulation algorithms, and greatly increased computer power [6]. The role of computer simulation is now well recognized in many fields e.g. drug design and toxicology [7–9]. In the same way, one can attempt to predict the detrimental effect of NPs from physical considerations. Establishing a qualitative and quantitative connection between physicochemical properties of the NP and their effect on biological functioning of membranes can help to identify the possible pathways leading to toxicity and give a mechanistic interpretation of toxicological data. To achieve this goal, one has to understand the processes occurring at bionano interface or on the initial stages of contact between the foreign nanomaterial and the organism such as formation of NP-biomolecule complexes, NP-cell membrane interaction, and NP uptake into the cell.

Understanding the corona formation and NP uptake requires one to address the lengthscales at the range of up to hundred nanometres, which is currently beyond the reach of direct atomistic modelling. Though lipid membranes have been very intensively studied by molecular simulations during last decade [10], in general, modelling NP translocation through a lipid membrane is a significant challenge. Depending on the size of the nanoparticle and any associated proteins (corona) tens of thousands, or more, lipid and other molecules may be needed to model a representative fraction of the membrane. For small (under 5 nm) NPs, cytotoxicity effects such as membrane disruption and poration can be addressed at the atomistic scale and at this scale significant insights have already been gained using molecular simulation using atomistic or coarse-grained force fields [11–14]. To assess interactions of larger NPs with membranes mesoscopic simulations based on greatly reduced number of degrees of freedom are required. To build a quantitative mesoscale model, information on NP-biomolecule association should be transferred from atomistic simulations to the mesoscopic scale.

Many of today's coarse-grained (CG) models use empirical parameterisation of effective interaction potentials. There exist several basic approaches for systematically constructing effective CG potentials from the results of atomistic simulations. One common approach is based on reproduction of forces for specific snapshots of the system (the force matching approach [15, 16] and the other, is based on fitting of structural properties, for which the radial distribution functions are typically used (the inverse Monte Carlo (IMC) or Newton Inversion method) [17, 18]. The IMC method was used to build coarse-grained models of various molecular systems including ion-DNA solution [19]. In the same spirit, CG models of plasma and membrane proteins have been developed [20–23] using the united-atom scheme, i.e. replacing the common groups of atoms by single beads, and thus drastically reducing the number of degrees of freedom. A systematic coarse-graining based on the all-atom presentations will preserve the shape and size of the relevant molecules and thus molecular specificity. In this approach we sacrifice a number of internal degrees of

freedom, such as protein conformations, which can be justified a posteriori. Although neglecting the protein internal degrees of freedom is a necessarily shaky approximation, this could be the most beneficial one as we can get around the dynamic bottlenecks related to slow protein unfolding.

Similar to molecules, one can use IMC and other coarse-graining methods to model effective interactions between nanoparticles [9, 24, 25]. Thus, the construction of the mesoscale modelling tool involves the following steps, with each consecutive stage based on a systematic coarse-graining of the more detailed description and validated by experimental data:

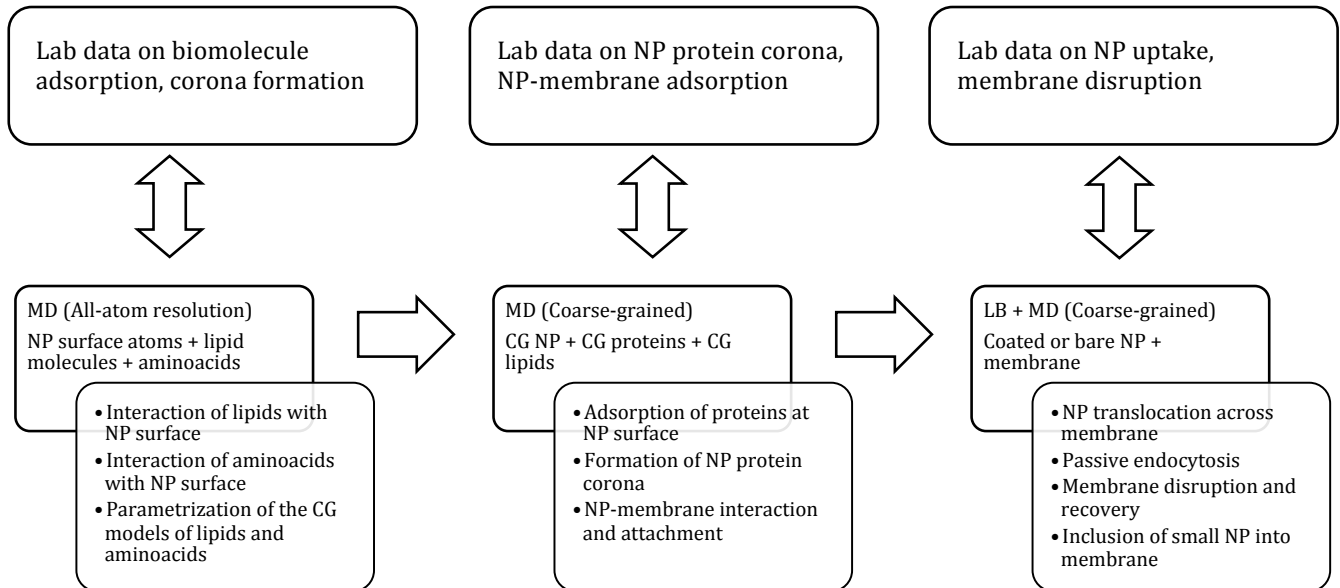


FIG. 1: Scheme of the multiscale simulation approach for modelling NP uptake.

In the following sections, we describe a set of CG tools that allow one to simulate the uptake of the NP-protein corona complex by a lipid bilayer. The remainder of the paper is constructed as follow. First we report a CG model to calculate the adsorption energies and the most favorable adsorption orientations of proteins onto a hydrophobic NP. The proposed method is then used to calculate the adsorption energies of the two common proteins in human blood onto NPs with negative or positive surface charge or neutral surface. We also report the effect of the NP radius on the adsorption energies and validate the proposed methodology against full atomistic simulations. Then, in Sec. III we describe a methodology in which full atomistic simulations of a lipid bilayer and various lipid-cholesterol mixtures are used for the extraction CG pair potentials. We also compare and validate the predictions of simulations at molecular and CG level. In Sec. IV, we present a CG simulation of the interaction a bare NP and of a NP-protein complex with a lipid bilayer. Finally, in Sec. V we summarise the main results.

## II. NANOPARTICLE-PROTEIN INTERACTION

It is now well accepted that foreign surfaces are modified by the adsorption of biomolecules such as proteins or lipids in a biological environment, and that cellular responses to materials in a biological medium might reflect the adsorbed biomolecule layer, rather than the material itself [26]. Recently, the concept of the NP-protein corona has been introduced to describe the proteins in association with NPs in biological fluids [27–30]. The composition of NP corona is flexible and is determined by many affinity constants and concentrations of the components of the blood plasma. One can speculate that in many practically relevant situations, the protein corona is the surface that is exposed to the cell membrane and is the entity the cell protective mechanisms have to deal with. Thus, for most cases it is more likely that the biologically relevant unit is not the particle, but a nanoobject of specified size, shape, and protein corona structure. Naked particle surfaces will have a much greater (non-specific) affinity for the cell surface than a particle hiding behind a corona of “bystander” proteins - that is proteins for which no suitable cellular recognition machinery exists. The evidence suggests that, in comparison to typical cell-membrane-biology event timescales the particle corona is likely to be a defining property of the particle in its interactions with the cell surface, whether it activates cellular machinery or not. Similar observations and outcomes exist for particles inside

Protein	PDB ID	Abbreviation	Weight fraction in plasma, %	Molar mass in, kDa
Human Serum Albumin	1N5U	HSA	5.0	67
Fibrinogen	3GHG	Fib	0.4	340

TABLE I: Proteins, PDB ID used for the coarse graining and the abbreviations used in the text, and their size and abundance in human blood plasma.

the cell, in key locations, though we cannot discuss details here [27–30]. We assume that the actual content of the corona is determined by (i) the NP exposure to the protein solution (blood plasma), (ii) a competition between the adsorbed proteins and the glycoproteins/membrane lipids. We will model the protein and lipid interaction with the NP surface at the CG united-atom level for selected set of proteins and lipids (see Table I). These simulations will provide interaction energies and will be used to predict the kinetics of protein/lipid corona formation. The data on aminoacid interaction with NP will allow us to compute binding affinities of arbitrary proteins of known structure within an additive model implying that the total protein-NP interaction energy is computed as a sum of NP interactions with aminoacids in contact with NP surface. This assumption will be further verified by comparison with results of large scale molecular dynamics simulations for a few selected protein types and mean force potential calculations. From the typical protein concentrations and adsorption energies one can also predict the average content of the corona using ideal adsorbed solution theory [31]. It is important to understand that at this stage we would be able neither to scan all the plasma proteins nor to take into account any change of protein conformation or bonding between the adsorbed proteins. However, as the effect of the corona is still largely unknown, we can only hope to capture the most important contributions of the plasma protein to the NP dispersion stability and their interaction with the cell membrane.

Due to complexity of blood plasma, we can only model it at a simplified level. It seems reasonable to include the elements, which are more likely to affect the NP interactions and aggregation, and mediate their interaction with the membrane. The plasma can then be modelled a solution of biomolecules in an implicit solvent with a dielectric constant of water and the Debye length corresponding to physiological ionic strength, van der Waals interactions set to corresponding triplets NP-protein-water, or protein-water-protein, and appropriate surface charges on the molecules. In this work we study the adsorption of two of the most abundant proteins in blood plasma, Human Serum Albumin (HSA) and Fibrinogen (Fib). In Table I, we summarize their relative content in blood and their molar mass. Although these two proteins represent important components of the blood plasma because of their abundance, recent observations [27–29, 32] demonstrate that the protein corona can include hundreds of different plasma proteins. As of now, it is not known which proteins dominate the content of the corona or play the most crucial role in the NP coating and uptake.

### A. Adsorption of proteins onto nanoparticles

The starting point for development of a CG model for the interaction of NPs with proteins is to decide how much detail from the molecular structure of the protein one needs to keep. There is an active and an extensive research activity on the different CG models that can be used to simulate proteins under different conditions and for detail reviews see [20, 22, 33]. The aim of this work is to propose a set of tools that could be used to simulate the interaction of one or more proteins (and in some cases quite big proteins) with a NP, for relatively long timescales. To meet this goal with a reasonable computational effort, the number of beads representing the protein should be as small as possible but the proposed model should also preserve enough structural information about the molecule. For these reasons we propose a single bead per residue model and consider the structure of the protein as a rigid body. We have studied the predictions of this model in more detail in Ref. [34]. The crystal structures of the proteins are obtained from the literature and one bead per amino-acid is placed at the position of the  $\alpha$ -carbon. At the end of this section we will test the validity of this first approximation. The second approximation is what level of detail will be needed to represent the NP. In this work, we will consider spherical homogeneous NPs so a single bead representation is justified.

In our model, the total NP surface-protein potential interaction ( $U$ ) is a function of distance of the surface to the centre of mass (COM) of the protein,  $d_{\text{COM}}$  and of protein orientation. It is given by a sum of two contributions:

$$U = \sum_i^N (U_i^{\text{VdW}} + U_i^{\text{el}}) \quad (1)$$

where  $N$  is the total number of residues in the protein,  $U_i^{\text{vdW}}$  is the van der Waals interaction of residue  $i$  with the surface and  $U_i^{\text{el}}$  is the electrostatics interaction of residue  $i$  with the surface.

For van der Waals contribution to the potential energy we propose a modified version of the residue-residue interaction potential as suggested in [21]. The model is based on the widely used residue-residue interaction energies proposed by Miyazawa and Jernigan [35], but instead of having a  $20 \times 20$  interaction matrix this is reduced to a table of normalized hydrophobicities,  $\epsilon_i$ , one for each amino acid (see Table II in [21]). A hydrophobicity index 0 is assigned to the most hydrophilic residue (LYS) and an index 1 to the most hydrophobic one (LEU). We should stress that any other hydrophobicity scale can also be used, it just has to be transformed such that the indexes have to be between 0 and 1, where 0 is assigned to the most hydrophilic residue while 1 to the most hydrophobic one. In this work, we consider a general surface which chemical reactivity that can be modeled as another residue with a hydrophobicity index  $\epsilon_s$ .

To model interaction of biomolecules with particles of different sizes, we use the following model for the nanomaterial. We assume that the interaction between a residue  $i$  and a bead of the nanomaterial  $s$  being at a distance  $r$  from each other is given by a modified 12-6 Lennard-Jones potential:

$$U_{s,i}(r) = \begin{cases} 4\epsilon_{e,n} \left[ \left( \frac{\sigma_{s,i}}{r} \right)^{12} - \left( \frac{\sigma_{s,i}}{r} \right)^6 \right] + \epsilon_{e,n}(1 - \epsilon_{s,i}) & r < r_{c,i} \\ 4\epsilon_{e,n}\epsilon_{s,i} \left[ \left( \frac{\sigma_{s,i}}{r} \right)^{12} - \left( \frac{\sigma_{s,i}}{r} \right)^6 \right] & r_{c,i} \leq r \leq r_{\text{cut}} \\ 0 & r > r_{\text{cut}} \end{cases} \quad (2)$$

$\epsilon_{e,n}$  is a parameter that scales the interaction energy,  $\epsilon_{s,i}$  is the combined hydrophobicity index of residue  $i$  and the nanomaterial and is given by  $\epsilon_{i,s} = \sqrt{\epsilon_i \epsilon_s}$ ,  $\sigma_{s,i}$  is the average van der Waals radius of residue  $i$  and the nanomaterial bead,  $\sigma_{s,i} = (\sigma_s + \sigma_i)/2$ ,  $r_{c,i}$  is the position of the minimum of the pair potential.

An integration of the 12-6 potential over the volume of the nanomaterial as defined in [21] gives a 9-3 Lennard-Jones-type potential. For a flat surface, the interaction can be expressed in terms of  $d$ , the distance between the residue centre of mass the closest element of the surface. An integration over a semi-space gives:

$$U_{si}^{\text{vdW}}(d) = \begin{cases} \epsilon_{es}\rho\sigma_{s,i}^3 \left[ \left( \frac{\sigma_{s,i}}{d} \right)^9 - \frac{15}{2} \left( \frac{\sigma_{s,i}}{d} \right)^3 + \left( \frac{125}{2} \right)^{\frac{1}{2}} (1 - \epsilon_{s,i}) \right] & d < d_{c,i} \\ \epsilon_{es}\epsilon_{s,i}\rho\sigma_{s,i}^3 \left[ \left( \frac{\sigma_{s,i}}{d} \right)^9 - \frac{15}{2} \left( \frac{\sigma_{s,i}}{d} \right)^3 \right] & d_{c,i} \leq d \leq d_{\text{cut}} \\ 0 & d > d_{\text{cut}} \end{cases} \quad (3)$$

where  $\epsilon_{es} = \frac{4\pi}{45}\epsilon_{e,n}$ ,  $\rho$  is the number density of beads in the nanomaterial,  $d$  is the distance from the residue  $i$  to the surface,  $d_{c,i} = (2/5)^{1/6}\sigma_{s,i}$ . Although the density  $\rho$  seems to be an important parameter scaling the interaction, it is not independent and therefore is not crucial for our method. From fitting the adsorption energy to experimental or MD simulation data, we can find the composite quantity  $\epsilon_{es}\rho$ , which is sufficient for further calculations. For a nanoparticle of radius  $R$ , a similar integration over the particle volume gives:

$$U_{si}^{\text{vdW}}(r) = \begin{cases} 4\epsilon_{es}\rho\sigma_{s,i}^3 \left[ \frac{(15r^6R^3 + 63r^4R^5 + 45r^2R^7 + 5R^9)\sigma_{s,i}^9}{(r^2 - R^2)^9} - \frac{15R^3\sigma_{s,i}^6}{(r^2 - R^2)^3} \right] - U_c^{\text{vdW}}(1 - \epsilon_{s,i}) & r < r_{c,i} \\ 4\epsilon_{es}\epsilon_{s,i}\rho\sigma_{s,i}^3 \left[ \frac{(15r^6R^3 + 63r^4R^5 + 45r^2R^7 + 5R^9)\sigma_{s,i}^9}{(r^2 - R^2)^9} - \frac{15R^3\sigma_{s,i}^6}{(r^2 - R^2)^3} \right] & r_{c,i} \leq r \leq r_{\text{cut}} \\ 0 & r > r_{\text{cut}} \end{cases} \quad (4)$$

where  $r$  is the distance from residue  $i$  to the centre of the nanoparticle. The distance  $r_{c,i}$  corresponds to the minimum of the potential and  $U_c^{\text{vdW}}$  is the value of the function  $U_{s,i}^{\text{vdW}}(r_{c,i})$  as defined in the range  $r_{c,i} \leq r \leq r_{\text{cut}}$ . We do not show the general expression for the position of the minimum as it is too bulky. The minimum is located at  $r_{c,i} - R \approx (2/5)^{1/6}\sigma_{s,i}$  at  $R \gg \sigma_{s,i}$  and is displaced to shorter distances at smaller  $R$ . The variation, however, is not very large, at  $R \rightarrow \infty$ ,  $r_{c,i} - R \approx 0.858374\sigma_{s,i}$ , at  $R = 200\sigma_{s,i}$  it is  $0.858375\sigma_{s,i}$ , at  $R = 20\sigma_{s,i}$  it is  $0.858469\sigma_{s,i}$ , at  $R = 2\sigma_{s,i}$  it is  $0.865242\sigma_{s,i}$ .

Note that the potential in Eq. (2) will only give a repulsive interaction between a highly hydrophilic surface and any residue (*i.e.* defining  $\epsilon_s = 0$ , gives  $\epsilon_{s,i} = 0$  for all residues). On the other hand, assigning a non-zero value for  $\epsilon_s$  will only change the magnitude of the interaction between any residue and the surface but not the shape of the potential. In this way, the proposed potential is limited to model only hydrophobic surfaces. Because of this limitation, we set the value of  $\epsilon_s = 1$  for all simulations. Alternatively, a potential that includes desolvation penalties, as the 12-10-6 Lennard-Jones potential proposed in [36, 37] for residue-residue interactions or the modified version proposed in [23] used to model residue-surface interactions, can be used to generate a more general interaction potential. The main drawback of the use of these more refined formulas for the potential is that the parameterisation is more challenging, and the applicability of a set of parameters could be very narrow.

Residue	LYS	GYU	ASP	ASN	SER	ARG	GLU	PRO	THR	GLY
$\epsilon_i$ ( $\mathcal{E}$ )	0.00	0.05	0.06	0.10	0.11	0.13	0.13	0.14	0.16	0.17
$\sigma_i$ (nm)	0.64	0.59	0.56	0.57	0.52	0.66	0.60	0.56	0.56	0.45
Residue	HIS	ALA	TYR	CYS	TRP	VAL	MET	ILE	PHE	LEU
$\epsilon_i$ ( $\mathcal{E}$ )	0.25	0.26	0.49	0.54	0.64	0.65	0.67	0.84	0.97	1.00
$\sigma_i$ (nm)	0.61	0.50	0.65	0.55	0.68	0.59	0.62	0.62	0.64	0.62

TABLE II: Normalized hydrophobicities  $\epsilon_i$  (taken from Table II in [21]) and  $\sigma_i$  for each amino acid. The most hydrophilic residue has a  $\epsilon_i$  of 0, while the most hydrophobic has a value of 1. For aminoacid-aminoacid interactions, we use the Lorentz-Berthelot mixing rules  $\sigma_{i,j} = (\sigma_i + \sigma_j)/2$ ,  $\epsilon_{i,j} = \sqrt{\epsilon_i \epsilon_j}$ .

The electrostatic interactions in Eq. (1) is modeled by adding point charges on the NP surface. This charges interact with the charged residues via a Debye-Hückel potential. The electrostatic interaction energy between a residue  $i$  and all the charges on the surface is given by:

$$U_i^{el} = \sum_j^{N_e} \lambda_B k_B T q_i q_j \frac{\exp(-r_{ij}/\lambda_D)}{r_{ij}} \quad (5)$$

where  $r_{ij}$  is the distance between the residue  $i$  and the point charge on the surface  $j$ ,  $\lambda_B = e^2 / (4\pi\epsilon_0\epsilon_r k_B T)$  is the Bjerrum length,  $k_B$  is the Boltzmann constant,  $T$  the temperature,  $\epsilon_0$  the dielectric permittivity of vacuum,  $\epsilon_r$  the relative dielectric permittivity of water,  $q_i$  the charge of residue  $i$ ,  $q_j$  the charge of the point charge  $j$  on the surface,  $N_e$  the total number of point charges on the surface and  $\lambda_D$  is the Debye length (defined through  $\lambda_D^{-2} = 8\pi\eta\lambda_B c_0$ , with  $c_0$  is the background electrolyte concentration). In practice, the points charges are evenly distributed on the spherical surface of the NP using a Golden Section spiral algorithm and all points will have the same charge  $q_j$  given by  $q_j = 4\pi\sigma R^2 / N_e$ , where  $\sigma$  is the surface charge density of the NP and  $R$  is the radius of the NP.

## B. Orientational sampling and the calculation of the adsorption energy

As we are assuming that the proteins are rigid bodies, we are not considering conformational changes during the adsorption process. Although, the adsorption process might conduce to conformational changes, this events happen at longer times than orientational changes on the surface [38]. Taking this into account, the adsorption energies calculated here will give an valuable insight into the long time formation of the NP-protein corona formation.

In our CG model, each residue of a protein is represented by a single bead located at the  $\alpha$ -carbon position. The native structures are obtained from the Protein Data Bank, and in Table I we report the proteins studied in this work, the PDB ID from which the CG model were built and the abbreviation that will be used in the rest of the text. The chosen proteins are some of the most abundant in human blood and will have a major influence in the formation of the NP protein-corona.

To identify the most favourable orientation of adsorbed protein globule (the one with the minimum adsorption enthalpy) we will follow the method suggested in [39], which is not as efficient as *e.g.* a genetic algorithm, but can provide additional information about the adsorption process. Briefly, a configuration space search is performed, where a systematic rotation of the protein allows us to build an adsorption map. There are three degrees of freedom (DOF) that have to be scanned. Fig. 2 shows that any point on the surface of the protein can be defined by a position vector from the COM of the protein. This vector is characterized by two angles:  $\phi$  and  $\theta$  and by rotating the molecule an angle  $-\phi$  around the  $z$  direction and then by an angle  $-\theta + 180^\circ$  around the  $y$  axes will make the position vector point towards the surface. The third DOF is the distance from the COM to the closest point of the surface,  $d_{\text{COM}}$ . Here, we sample  $\phi$  from 0 to  $350^\circ$  in steps of  $10^\circ$  and  $\theta$  from 0 to  $170^\circ$  in steps of  $10^\circ$  (note that  $\phi = 0^\circ$  is equivalent to  $\phi = 360^\circ$ , and that  $\theta = 0^\circ$  is equivalent to  $\theta = 180^\circ$ ). Instead of obtaining the “real” adsorption free energy by calculating the potential of mean force for all orientations, we only calculate the potential energy  $U$  (given by Eq. (1)), which is the sum of all the interactions between the surface and the protein. As the adsorption energies are expected to be at least 5 times  $k_B T$  and as the proteins are assumed to be rigid, neglecting thermal fluctuations is clearly justified. For each configuration  $(\phi_i, \theta_j)$ , the total potential energy is calculated as a function of distance of the COM,  $U(d_{\text{COM}}, \phi_i, \theta_j)$ , to the surface for the case of a slab (Fig. 2b) or to the center of the NP for the case of a NP (Fig. 2c). Following a similar approach as in [40], and denoting the reaction coordinate  $d_{\text{COM}} = z$ , the adsorption

energy for any particular configuration in the case of a protein adsorbing on a flat surface is given by:

$$E(\phi_i, \theta_j) = -k_B T \ln \left[ \frac{1}{a(\phi_i, \theta_j)} \int_0^{a(\phi_i, \theta_j)} \exp(-U(\phi_i, \theta_j, z)/k_B T) dz \right] \quad (6)$$

where  $a(\phi_i, \theta_j)$  is the maximum interaction distance from the COM of the protein to the surface for the given orientation. For the case of a NP-protein interaction, the mean interaction energy for any particular orientation is given by:

$$E(\phi_i, \theta_j) = -k_B T \times \ln \left[ \frac{3}{(R+a(\phi_i, \theta_j))^3 - R^3} \int_R^{R+a(\phi_i, \theta_j)} z^2 \exp\left(\frac{-U(z, \phi_i, \theta_j)}{k_B T}\right) dz \right] \quad (7)$$

Then the total mean adsorption energy of the system for both cases (slab and NP),  $E_{ad}$ , can be estimated by averaging over all adsorbed states with Boltzmann weighting [39]:

$$E_{ad} = \frac{\sum_i \sum_j P_{ij} E(\phi_i, \theta_j)}{\sum_i \sum_j P_{ij}} \quad (8)$$

where  $P_{ij} = \sin(\theta_j) \exp[-E(\phi_i, \theta_j)/k_B T]$  is the Boltzmann weighting factor.

### C. Details of the simulations, parametrization and validation

All simulation were performed using Espresso [41] and the cutoff for the interaction potential in Eq. (2) was set to  $r_{cut} = 6$  nm. For all calculation the simulation box was taken big enough to fit the NP and the protein. The method described here only involves the calculation of the total energy of the system given by Eq. 1, therefore a coupling to a thermostat is not required. After the CG model were built from the PDB files, the obtained structures were shifted so the COM of the molecules was in the origin of the frame of reference and this structure was defined as the  $(\phi = 0^\circ, \theta = 0^\circ)$  orientation. With this definition the first residue in the sequence of each protein will have the following  $(\phi, \theta)$  angles:  $(21.4^\circ, 85.2^\circ)$  for HSA and  $(132.1^\circ, 46.4^\circ)$  for Fib.

The units of the simulations are: lengths ( $\mathcal{L}$ ) in nm, energy ( $\mathcal{E}$ ) in  $k_B T \approx 4.15 \times 10^{-21}$  J taking a temperature of  $T = 300$  K, for the mass unit ( $\mathcal{M}$ ) we selected the average mass of the 20 residues (ca. 110 Da) hence in our simulations all residues have a mass of 1. The values of  $\epsilon_i$  and  $\sigma_i$  can be found in Table II and as mentioned in Sec. II A we will only consider hydrophobic NPs with  $\epsilon_s = 1$  and  $\sigma_s = 0.35$  nm.

NPs with negative surface charges as well as neutral NPs were considered. For the negative charged cases, a surface charge density of  $-0.02$  C/m<sup>2</sup> was used. As explained in Sec. II A, the charged surfaces are modelled by individual point charges. The surface density of these charged beads ( $\sigma_c = N_e/R^2$ ) was set to  $4$  nm<sup>-2</sup> for all the simulations, which gives *e.g.* a  $N_e = 100$  for a NP of  $R = 5$  nm. Then, we assumed that each bead carries a charge of  $-0.39e$ , where  $e$  is the elementary charge. As we are considering physiological conditions, we use  $\lambda_B = 0.73$  nm and  $\lambda_D = 1$  nm. Residue charges at this condition are  $+e$  for LYS and ARG,  $-e$  for ASP and GLU, and  $+0.5e$  for HIS. The rest of the residues are neutral.

The only free parameters of the model are  $\rho\epsilon_{es}$  in Eq. (2), and the parameterisation was done by systematically changing its value to match experimental data of adsorption of Lysozyme on hydrophobic surfaces reported by Chen *et. al.* [42]. The native structure for our CG model of Lysozyme was obtained from the PDB ID:2LYZ. With  $\rho\epsilon_{es} = 1.972k_B T/nm^3$  we obtain a value of  $-7.6k_B T$  for the adsorption energy (very close to the experimental reported value of  $-7.9k_B T$ ).

To validate the parameterisation, the adsorption energy of Myoglobin (PDB ID: 1MBN used for the CG model) was calculated using the same value of  $\rho\epsilon_{es}$  obtained from the parameterisation. In this way, a value of  $-6.1k_B T$  was found for the adsorption energy of Myoglobin. This value is slightly lower than the experimental value of  $-7.6k_B T$  also reported by Chen *et. al.* [42] but reproduces the trend that Myoglobin adsorbs slightly weaker than Lysozyme to a hydrophobic surface.

### D. Protein adsorption energies

Results for the adsorption energies calculated using Eq. (8) as a function of NP radius are shown in Fig. 3. The results show that HSA adsorbs stronger as the radius of the NP increases until it reaches a minimum value (Fig. 3a). For small

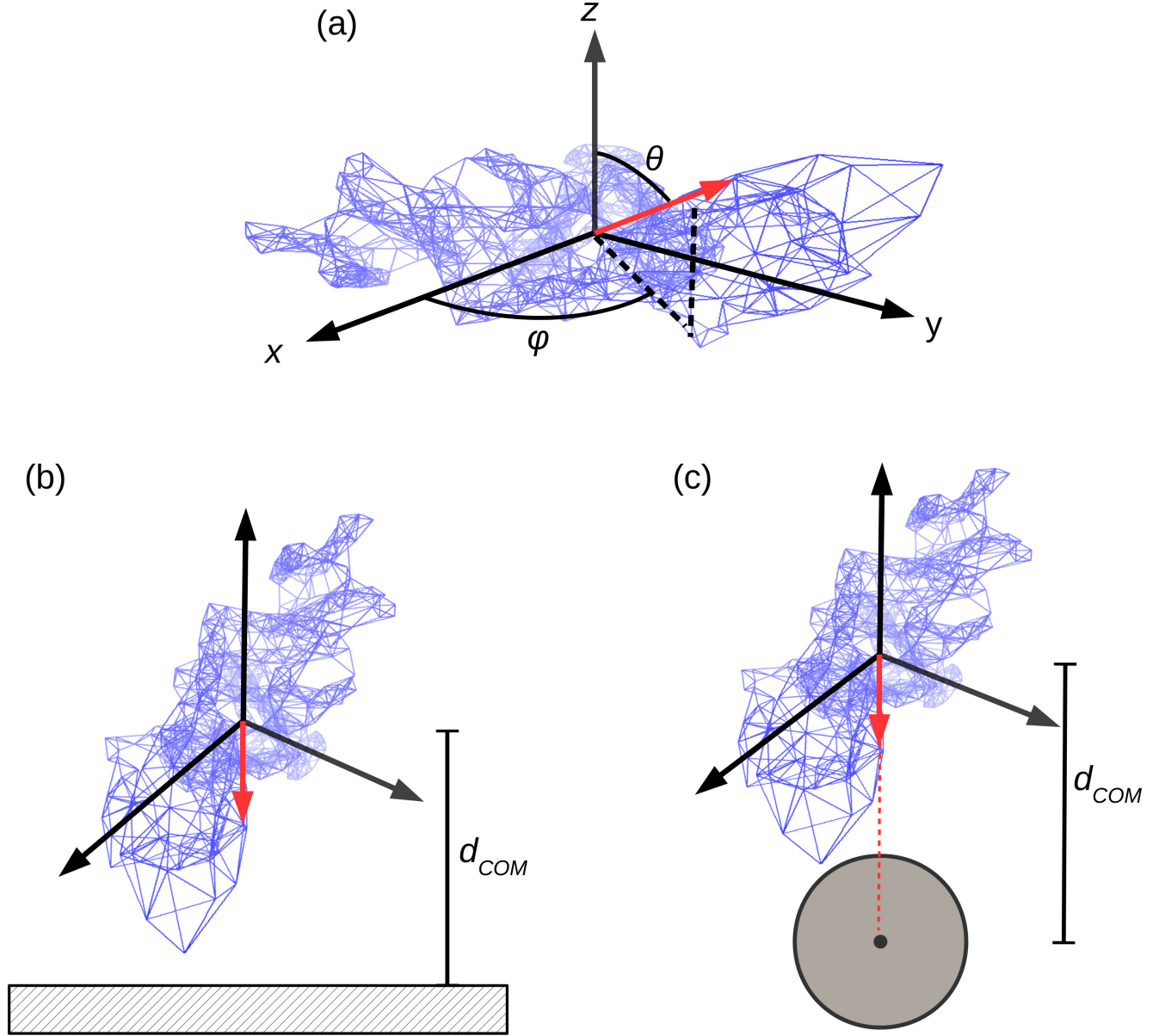


FIG. 2: Definition of the protein orientation. (a): Any point on the surface of the protein can be defined by a position vector from the COM to that point and depends on two angles  $\phi$  and  $\theta$ . The remaining degree of freedom is the distance of the COM,  $d_{COM}$  to (b) the surface for a slab or (c) to the center of the NP for a NP.

NPs, the combination of the curvature effect (increasing  $R$  increases the van der Waals interactions) with the availability of residues to interact with the surface originates that the proteins adsorb stronger (more negative values) as the radius is increased. Then after a value of radius around 50 nm the  $E_{ad}$  starts to converge to the value corresponding to a flat surface as the van der Waals interactions and the number of residues close to surfaces do not change significantly by increasing  $R$ . We performed calculations for NPs of  $R$  up to 500 nm and confirmed that the adsorption energy indeed converges to the slab values. For the Fib molecule the situation is different (Fig. 3b). In this case the adsorption energy decreases as a function of  $R$  at least until the biggest radius studied here ( $R = 100$  nm) and it is lower than that for the adsorption onto a flat surface. The big size of the Fib molecule (ca. 45 nm on its longest axes) makes that for at least until  $R = 100$  nm the combined effects of curvature and number of residues that interact with the surface are still noticeable. The effect of the charge is more important for the HSA than for the Fib. HSA is overall negative, so the electrostatic interactions contribution is mainly repulsive

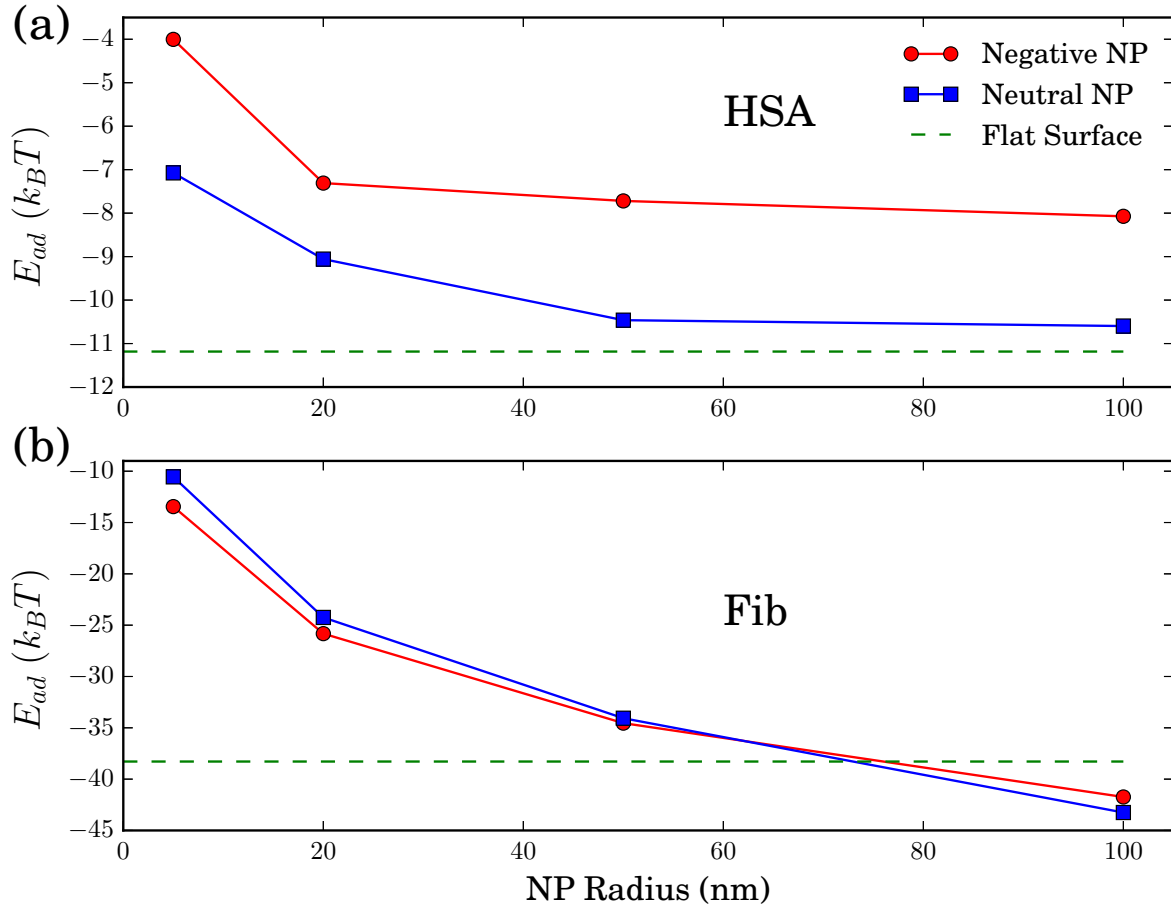


FIG. 3: Adsorption energies as a function of the NP radius for the proteins studied and the three types of surface charge: Negative:  $-0.02 \text{ C/m}^2$  and Neutral: no charge. (a) HSA and (b) Fib. The dashed lines show the Adsorption energy for the case of a flat surface.

increasing the values of the  $E_{ad}$ . On the other hand, the Fib molecule is positive and the electrostatics interactions tend to increase the adsorption of the Fib onto a negative surface. In neither of the proteins the effect of electrostatic interactions was more than  $3 k_B T$ .

The systematic sampling employed for the calculation of the adsorption energies can also be used to identify the most favourable orientations for adsorption and to study how the charge and/or the radius of the NP influence the protein orientation. Fig 4, shows a surface map of the adsorption energy as a function of the angles  $\theta$  and  $\phi$  for HSA. Each panel is for a radius of 5 or 100 nm and for a neutral or a negative charged surfaces. In general, the surfaces are complex in structure showing an energy landscape with several local minima with differences less than  $1 k_B T$ . It is also important to notice that the maps have large areas with adsorption energies of  $-6 k_B T$  or lower. Our results show that HSA will strongly adsorb at physiological conditions and room temperature and that orientational changes after adsorption are energetically favourable. Comparison of different panels in Fig 4 shows that radius has only a small effect on the preferred orientations, while the charge density has an unnoticeable impact on the preferred orientations.

A different scenario is observed for Fib. Fig 5 shows a surface maps of Fib adsorption for two radii for neutral and charged surfaces. In this case, the surface maps depend on the radius of the NP (compare Fig 5a with Fig 5b or Fig 5c with Fig 5d) but change very little between the charged and uncharged surface (compare Fig 5a with Fig 5c or Fig 5b with Fig 5d). As we already noticed for HSA, the charge has a small effect on the total adsorption energy so it is expected that it would not dramatically change the surface maps. The radius and the surface curvature seem to be more important as for big proteins (like Fib) a larger NP allows a more extensive contact and thus influences the preference for protein orientations (or NP binding pockets). In Fig. 6 we show the most favorable orientations for Fib



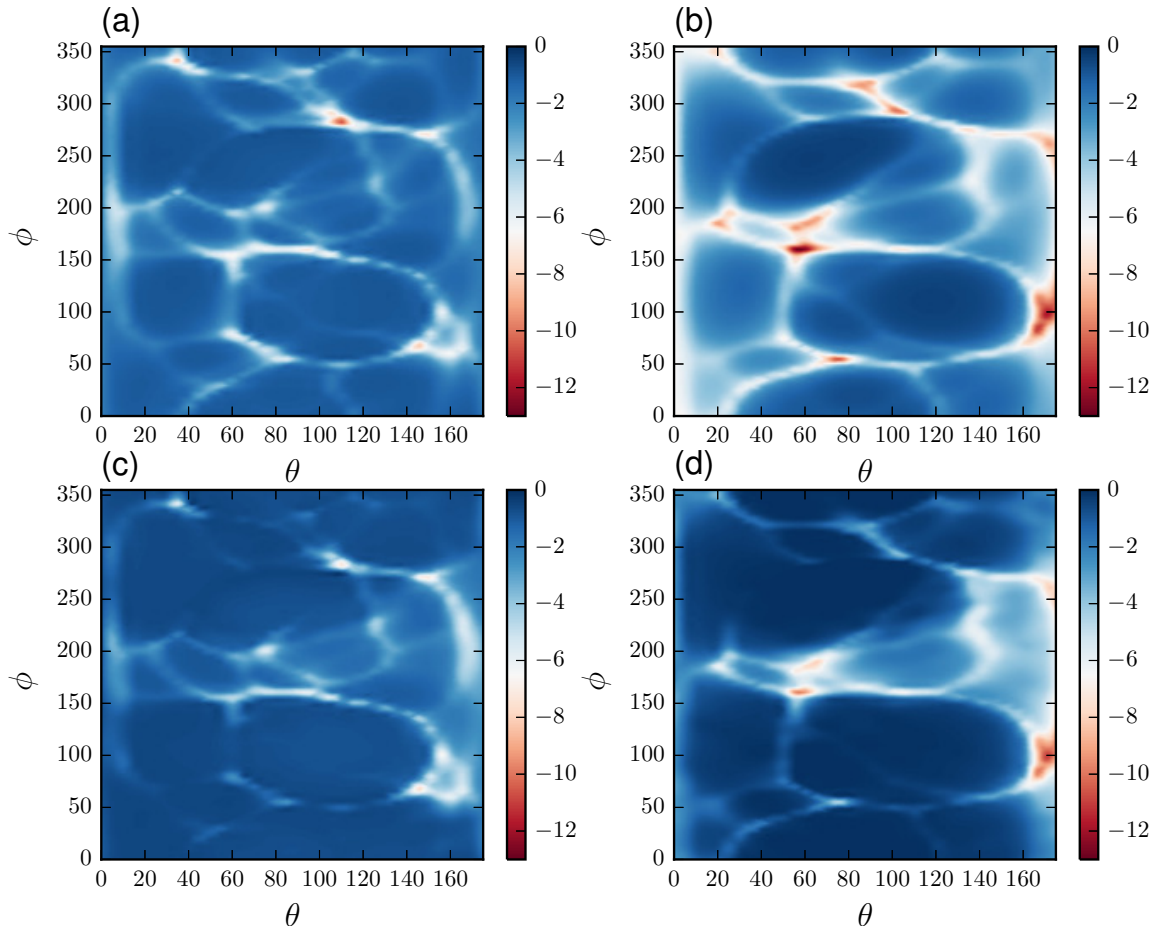


FIG. 4: Adsorption energy maps for HSA. (a) Neutral particle of  $R = 5$  nm. (b) Neutral particle of  $R = 100$  nm. (c) Negatively charged particle of  $R = 5$  nm. (d) Negatively charged particle of  $R = 100$  nm.

on a neutral surface for two different NP radii. For the small NP (Fig. 6a), Fib has its adsorption energy minimum in a configuration where the NP interacts with a relatively small segment of the molecule. Meanwhile for a large NP, Fib tends to adsorb in a completely different orientation (Fig. 6b). Now the most favourable orientation is one that oriented with the longest axis of the Fib molecule along the surface.

A very straightforward conclusion from the above data is that the bigger the protein, the stronger it will adsorb onto a NP. This result agrees with the experimental observation reported by De Paoli *et. al* [43], which show that the binding association constant on citrate coated gold NPs (which can be considered as moderately negative hydrophobic NP) depend mainly on the size of the protein (they studied HSA, Fib and other blood proteins). It is interesting also to compare our results with the simulations of NP corona formation reported by Vilaseca *et al.* [44]. Using CG MD simulations they found that for a flat surface at long times the most abundant protein adsorbed were Fib, then immunoglobulin- $\gamma$  (of intermediate size between HSA and Fib) and at last HSA. At this point, it is important to remark that the adsorption energies calculated in this work could be a good predictor of the equilibrium composition of the NP-protein corona but at short times other factors such as the protein sizes and their concentrations have to be considered to predict the corona composition.

### E. Validation of the methodology

We now test our CG methodology with predictions of full-atomistic MD simulation. We model adsorption of small plasma protein Ubiquitin (Ubi) to a flat  $\text{TiO}_2$  surface. The reasons of choosing Ubi for the validation was due to

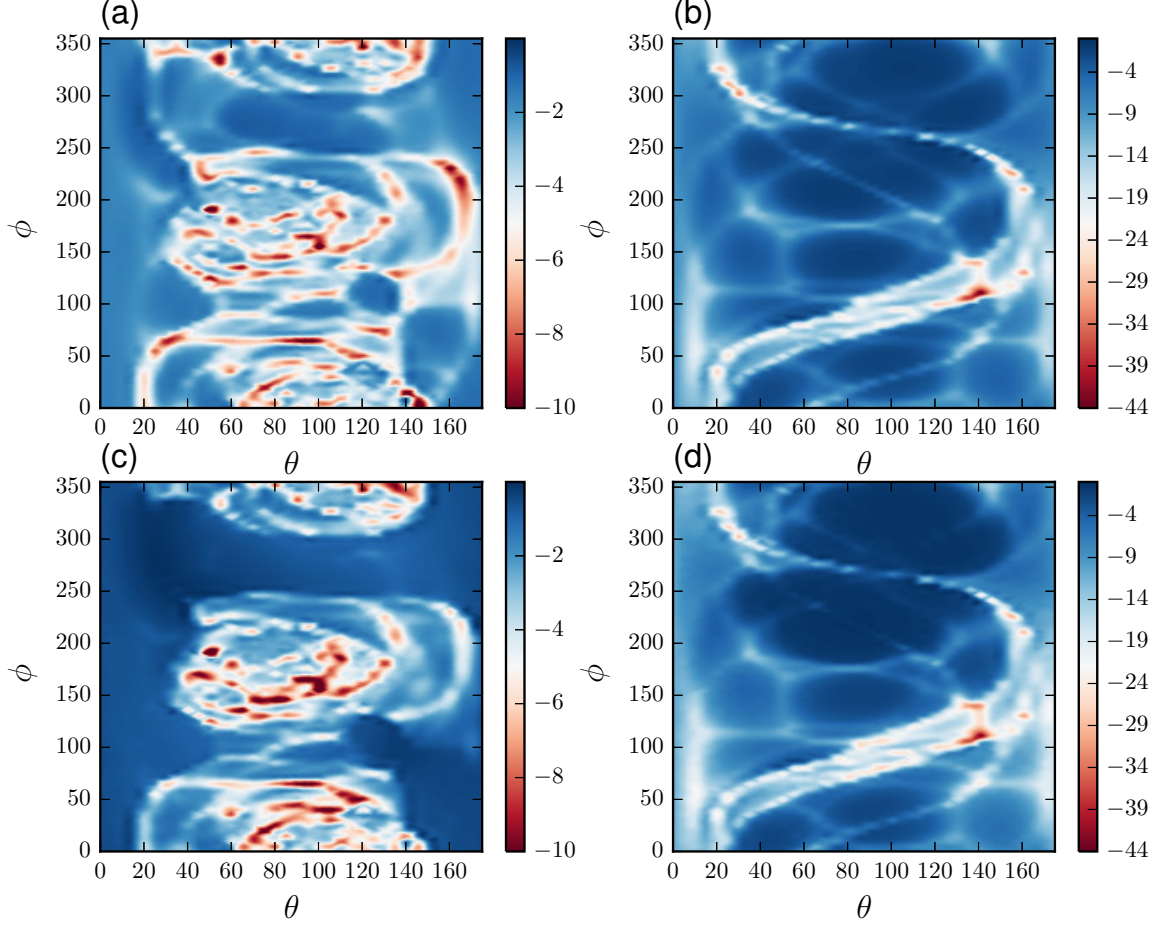


FIG. 5: Adsorption energy maps for Fib. (a) Neutral surface and  $R = 5$  nm. (b) Neutral surface and  $R = 100$  nm. (c) Negatively charged surface and  $R = 5$  nm. (d) Positively charged surface and  $R = 100$  nm.

the it size (only 76 residues) and known folded structure, which allows us to perform full atomistic simulations in a reasonable amount of time.

The Ubi crystal structure was obtained from the PDB (PDB ID file 1Ubi [45]) and was CG as explain above (see Fig. 7). To be able to directly compare against full atomistic simulations, in this case the potential interaction between the 20 different residues and the surface were obtain by performing full atomistic simulations of the adsorption of each of the 20 aminoacids and then performing an inverse Monte Carlo.

With the CG methodology we obtained the total adsorption energy of  $-10.7k_B T$  for Ubi and the adsorption map is shown in Fig. 8. The surface obtained shows two major minima and a number of local minima. It also predicts that orientational changes are favorable after adsorption as some of the minima are connected by an energy landscapes with rather small barriers (less than  $3k_B T$ ).

To analysis the validity of our model and to understand the dynamical behavior of the adsorption process we now preform a series full atomistic simulations. We used the VESTA program [46] to construct a  $5 \times 20 \times 32$ -supercell of the  $\text{TiO}_2$  rutile unit cell. The coordinates were rotated so that the normal of the  $\text{TiO}_2$  slab, corresponding to the (100) surface, was oriented along the  $z$ -direction. The box was elongated in the  $z$ -direction and periodicities were kept in all directions. The final size of the simulation box was then  $9.466 \times 9.184 \times 12$  nm. Covalent bonds were added to all Ti-O pairs within a 2 Å-cutoff. We used force field parameters for the  $\text{TiO}_2$  slab from a recent parameterisation study [47]. This same force field was use to calculate the CG potential interactions between the surface and the 20 amino acids. The same folded Ubi structure as for the CG model was used and inserted above the  $\text{TiO}_2$  slab. The  $\text{TiO}_2$ -Ubi system was solvated by insertion of 25817 water molecules around the protein and the slab, and the final system contained 99802 atoms. The system was energy minimized for 1000 steps and then equilibrated at constant temperature (300

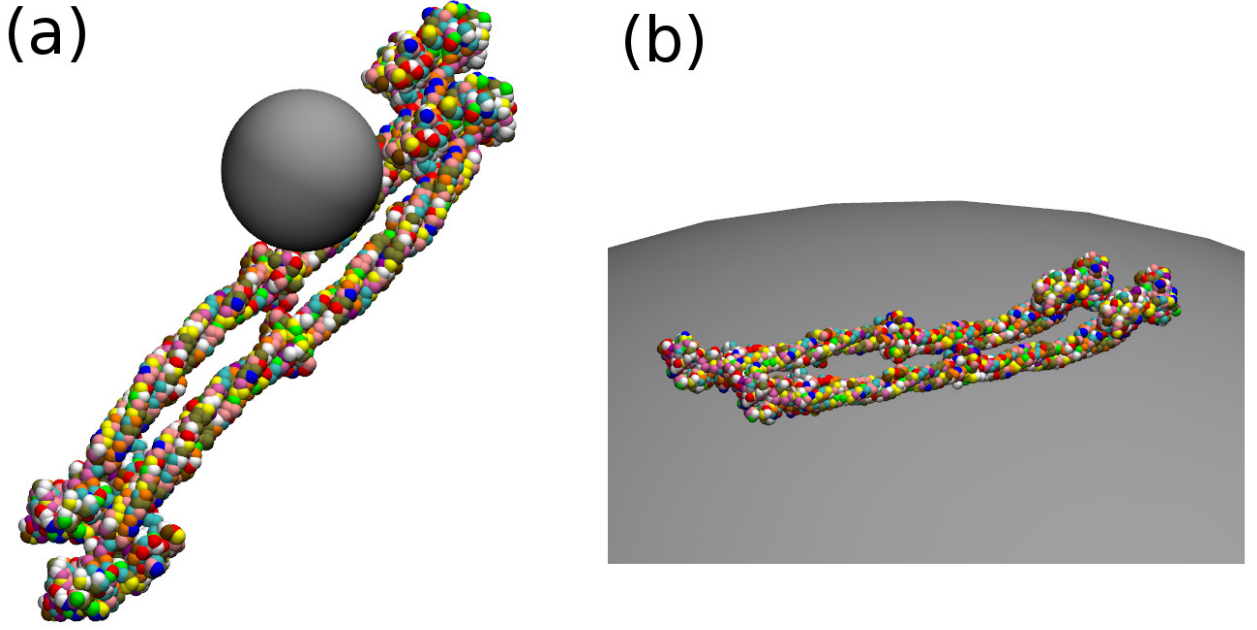


FIG. 6: Fib most favorable orientation for adsorption on a neutral surface: (a)  $R = 5$  nm and (b)  $R = 100$  nm.

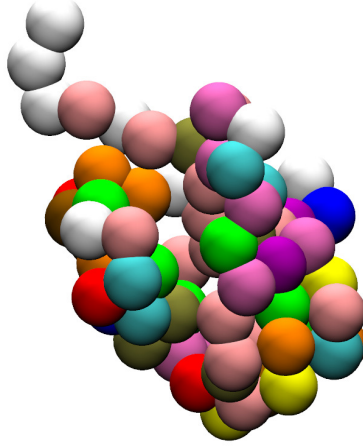


FIG. 7: CG model of Ubi (PDB ID: 1Ubi [45]). In our model each residue in the protein is represented by one bead

K) and pressure (1 bar) for 100 ps using Berendsen's weak scaling algorithms [48], with relaxation constant  $\tau = 1$  ps in both cases. The temperature coupling was applied independently to the  $\text{TiO}_2$  slab and to the rest of the system. The pressure tensor must be kept anisotropic due to the solid  $\text{TiO}_2$  slab, but the off-diagonal components of the compressibility tensor (and the reference pressure tensor) were set to zero to enforce a rectangular simulation box. The diagonal elements of the compressibility tensor were set to  $5 \times 10^{-7} \text{ bar}^{-1}$  in the lateral directions (bulk  $\text{TiO}_2$ ) and  $5 \times 10^{-5} \text{ bar}^{-1}$  in the normal direction (bulk water). The box vectors relaxed 2-4% during equilibration.

In a first simulation, we placed the protein in the  $(\phi = 0^\circ, \theta = 0^\circ)$  orientation close to the surface and followed the dynamics for 440 ns at constant volume and 300 K. The Nose-Hoover thermostat [49, 50] with the coupling constant

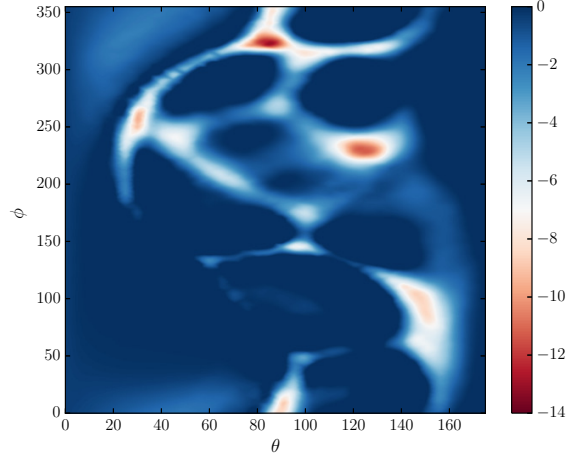


FIG. 8: Adsorption energy maps for Ubi adsorbing into a  $\text{TiO}_2$  slab.

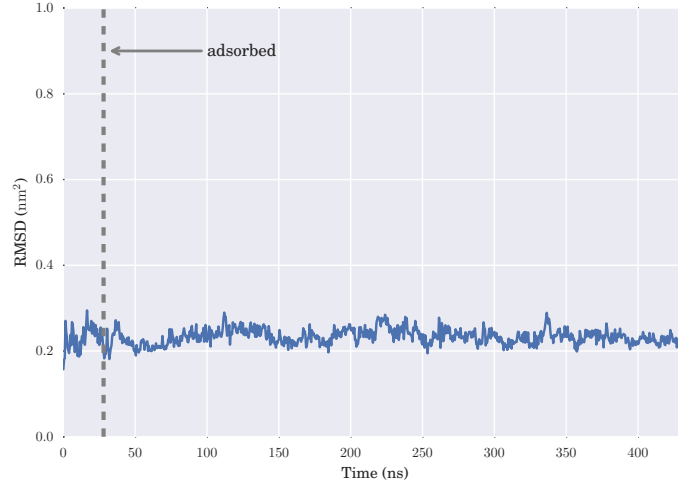


FIG. 9: The root-mean-square deviation (RMSD) of Ubi during the simulation with respect to the PDB reference structure. No unfolding occurs and the RMSD is  $0.15 \text{ nm}^2$  throughout the simulation, which is the same as found in simulations of the folded structure in bulk water.

$\tau = 5 \text{ ps}$  was used to ensure proper sampling of the ensemble when controlling the temperature. The simulation was run in parallel using 512 cores and frames were kept every 5 ps.

The trajectory obtained showed that the protein motion was diffusive in the bulk water for about 25 ns until making contact with the  $\text{TiO}_2$  slab. Then the Ubi molecule attached to the surface and remained adsorbed for the rest of the simulation. To study the stability of the structure of the protein during adsorption we calculate the root-mean-square-deviation (RMSD) as a function of time and the results are shown in Fig. 9. The RMSD remained at a low constant value of ca.  $0.2 \text{ nm}^2$  during the simulation, i.e., no unfolding occurred in the adsorbed state. This results clearly verifies that for the adsorption of Ubi on  $\text{TiO}_2$ , a rigid body model for the protein structure is well justified.

A detailed study of the simulation trajectory revealed that the protein motion could be characterized by four states, and that adsorption occurs through a two-step mechanism (Fig. 10). First, the protein diffuses freely in the bulk water. Second, the C-terminus of Ubi the protein contacts the  $\text{TiO}_2$  surface and provides a lock for the protein to the first solvation layer. Third, Ubi rotates and locks into position on the surface. Fourth, the protein diffuses on the surface in the locked orientation.

The adsorption mechanism is relatively fast once the first surface contact is initiated. The anchoring of the C-terminus to the first solvation layer occurs in about 5 ns and the locking is completed after 10 additional ns. The residues at the C-terminus are ARG 74, GLY 75 and GLY 76. The anchoring is initiated when the charged end of ARG74 contacts the surface perpendicularly. The contact is not with the bare surface, but with the first solvation

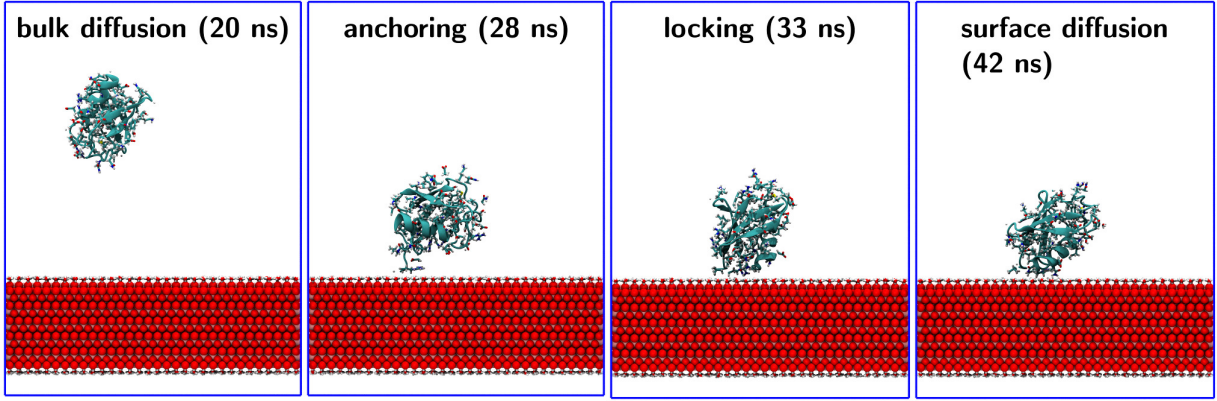


FIG. 10: Four distinct stages were identified during adsorption. (a) The protein diffuses in the bulk water. (b) The N-terminal of Ubi anchors to the solvation layer of the  $\text{TiO}_2$  slab. (c) The protein rotates and locks to the solvation layer through GLN40 and GLN31. (d) The protein diffuses on the solvation layer in the locked conformation.

layer, which is strongly bound directly to the surface. At this point the rest of the protein diffuses in the bulk until (ca. 5 ns) GLU 40 can contact the surface, which leads to the protein being locked into an adsorbed orientation after 10 ns. The locking procedure consists of Ubi first connecting GLN40 to the surface, followed by a second connection through GLN31. The two glutamines (GLN31 and GLN40) form a bridge that stabilize the orientation of the protein and no more change in orientation occur for the rest of the simulation.

The residues involved in the anchor-lock mechanism are arginine and glutamine. These have been identified by potential of mean force calculations [47] of isolated side chain fragments (together with aromatic side chains) to be the strongest binders to  $\text{TiO}_2$ . In both cases, the  $\text{NH}_2$ -group of the end of the amino acid approaches the surface in a perpendicular orientation, but can then rotate to maximize the interactions with the solvation layer.

The “upright” position of the protein in the adsorbed state suggests that it does not correspond to a free energy minimum. Since the orientation does not change over 400 ns, it is likely that there are high free energy barriers associated with the orientation changing into another free energy basin. To map all values of  $\phi$  and  $\theta$ , more sampling of the protein adsorption is needed. This could either be done in a repetitive fashion from different starting configurations or with enhanced sampling techniques such as metadynamics.

As for Ubiquitin we do not observe any unfolding within several hundreds of nanoseconds, we can conclude that our rigid protein model for studying adsorption is justified at least for some conditions: small nanoparticles, non-metallic particles and small and compact proteins such that the adsorption energies are within few tens of  $k_B T$ . For other situations, one should evaluate the energy to decide whether the model is sufficient. In general, conformational changes can be an important factor for the adsorption dynamical process [43]. This assumption can be relaxed by *e.g.* using a Gō-Type model (see [33] for a review on CG models of proteins). Furthermore, as the methodology we presented is computationally very efficient and can provide information about the structure of NP-protein complexes, it can be used as an exploring tool to perform more sophisticated and computationally demanding calculations.

### III. COARSE-GRAINED MODEL OF A LIPID BILAYER

Any attempt to simulate with some molecular details but at length and time scale involve in the uptake of NPs through a cell membrane must rely on a CG model of the main constituents of the biological membranes. In this section we describe a methodology to systematically CG a lipid bilayer and lipid bilayer containing cholesterol from the results of full atomistic simulations.

#### A. Molecular simulations of various lipid-cholesterol mixtures

We started the CG procedure by performing all-atom molecular dynamics simulations for three lipid mixtures: (i) 1,2-dimyristoyl-sn-glycero-3-phosphatidylcholine with cholesterol (DMPC + CHOL); (ii) 1,2-dioleoyl-sn-glycero-3-phosphatidylserine with cholesterol (DOPS + CHOL); (iii) 1,2-dimyristoyl-sn-glycero-3-phosphatidylcholine with 1,2-dioleoyl-sn-glycero-3-phosphatidylserine (DMPC + DOPS). The composition of these systems are reported in the Table III A. In each simulation, the starting state was generated randomly and was energy minimized afterwards. Then



System	I. DMPC-cholesterol	II. DMPC-DOPS	III. DOPS-cholesterol
Number of DMPC	30	30	-
Number of DPPS	-	30	30
Number of cholesterol	30	-	30
Number of water	2000	2000	2000
Number of Na+	-	30	30

TABLE III: Composition of the simulations used for the CG of lipids mixtures.

a short 1 ns NVT simulation at density 1 g/cm<sup>3</sup> have been carried out, which was followed by 100 ns equilibration simulation in NPT-ensemble and production stage of 400 ns. The Slipids force field was used [51, 52]. Other simulation parameters: time step 2 fs; Nose-Hoover isotropic thermo/barostat with temperature 303 K, pressure 1 bar, relaxation times 0.1 and 1 ps for thermostat and barostat respectively; all bonds were constrained by Links algorithm; particle-mesh Ewald with Fourier spacing 1 Å and tolerance parameter 10<sup>-5</sup>. The configurations were saved in the trajectory each 10 ps. The atomistic simulations were performed using the Gromacs simulation engine (v. 4.5) and a rigid TIP3P water model.

### B. Mapping of atomistic to coarse-grained trajectories – from residue to beads

The atomistic trajectories obtained in the simulations were mapped onto coarse-grained trajectories, and radial distribution functions between sites of the coarse grained models have been determined. As shown in Fig. 11, 10 beads for representation of DMPC molecule were used at the CG level (3 beads instead of each of the two hydrocarbon tails, 4 beads instead of the head group including esters), 14 beads for DOPS molecule (5 beads instead of each hydrocarbon tail with specific distinguishing of the beads with double bond and beads uniting 3 or 4 methylene groups, and 4 beads instead of the head group), 5 beads for CHOL molecule, and Na+ ions as a single bead were used. Water was not included into CG model but its effect was included into solvent-mediated potentials. Fig. 12 shows a snapshot of CG DMPC bilayer, which is spontaneously formed in a CG lipid system.

The radial distribution functions (RDF) between CG sites obtained after coarse-graining of the atomistic trajectories were used to compute effective potentials defining interactions in the coarse-grained models using the inverse Monte Carlo method. The RDF were computed for each pair of different coarse-grained sites and were used as an input to compute effective coarse-grained potentials which reproduce the RDFs. Computations of effective potentials were done for the same compositions of the systems I, II, and III as the respective atomistic simulations listed in Table III A. The software package MagiC [53] was used. In the inversion process, the first 20 iterations have been carried out using iterative Boltzmann inversion, followed by 30-40 iterations using the inverse Monte Carlo algorithm.

More specifically, the RDF's have been determined between beads involved in “non-bonded” interactions, that is between CG sites belonging different molecules or the same molecule but separated by more than two bonds. Also, reference distribution functions for the bond lengths and bending angle distribution functions were determined for the all CG sites relevant for the three types of considered molecules. Then the calculated RDFs and bonded reference distribution functions were used to calculate parameters of the corresponding coarse-grained potentials. This is a multistage process, from a high resolution system description to a low resolution one. Monte Carlo computer simulations of the coarse-grained system DMPC + CHOL using Metropolis method (MagiC package) were carried out. The parameters were calculated using a two step iteration technique: first, the iterative Boltzmann inversion method was performed to calculate a set of intermediate parameters; second, the inverse Monte Carlo algorithm was used to calculate the final set of parameters. The final parameters of the coarse-grained potentials for DMPC + CHOL mixture have been calculated.

### C. Validation of the lipid coarse-grained model

The interaction potentials obtained for the CG models using the inverse Monte Carlo technique were validated by comparison with atomistic simulations. Figures 14 show radial distribution functions between some selected sites of DMPC lipid and Cholesterol computed in coarse-grained and atomistic simulations of a mixture of 30 DMPC lipids, 30 cholesterol molecules and 1800 waters. The result shows a perfect coincidence of the RDFs, which justifies the approximations made and the quality of the CG model.

Fig. 12 shows a snapshot of CG DMPC bilayer, which is spontaneously formed in a CG lipid system. We have

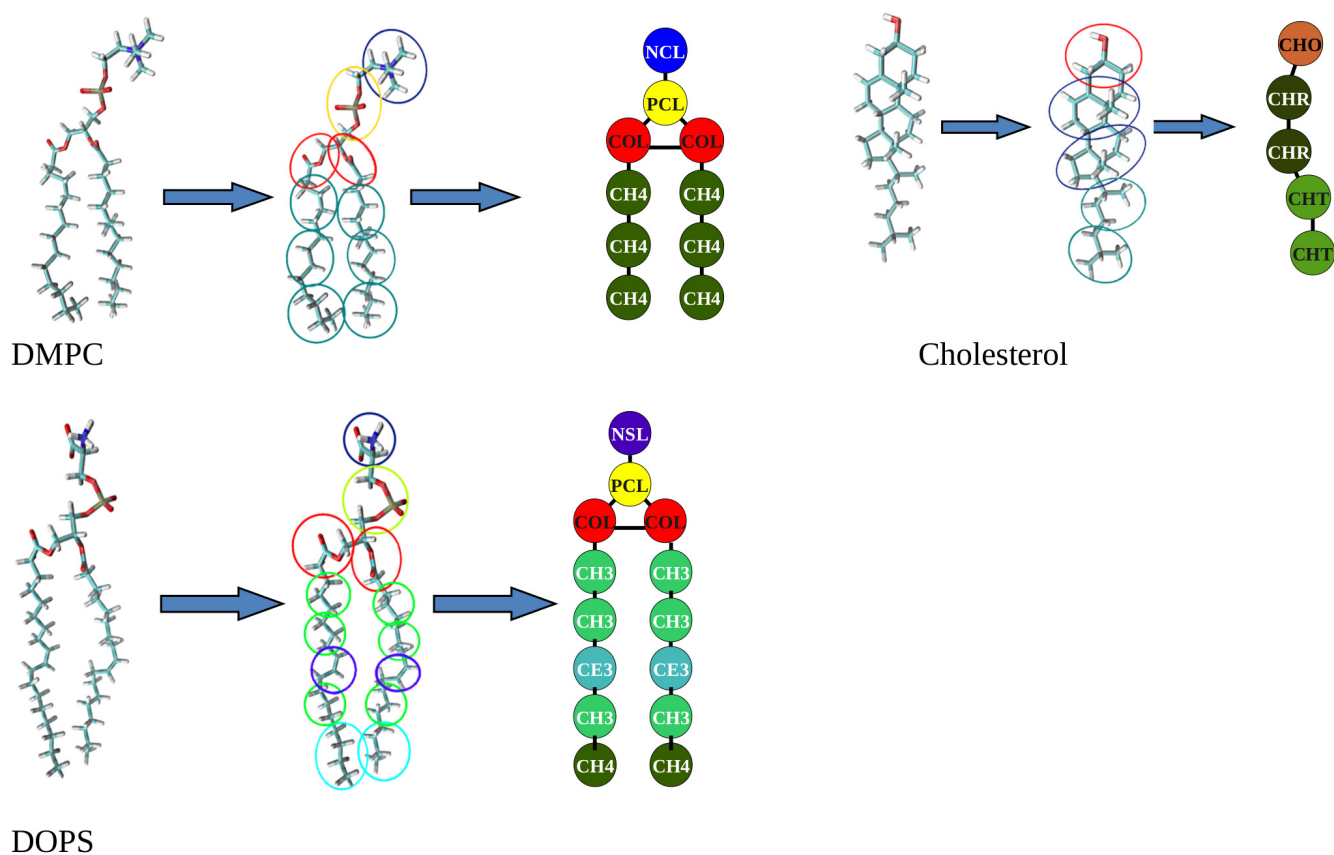


FIG. 11: Mapping of systems at an atomistic level to a coarse grained level where each residue of the atomistic system is replaced by a bead for DMPC, Cholesterol and DOPS (1,2-dioleoyl-sn-glycero-3-phosphatidylserin) molecules. Coarse-grained sites of the same type are given by the same color.

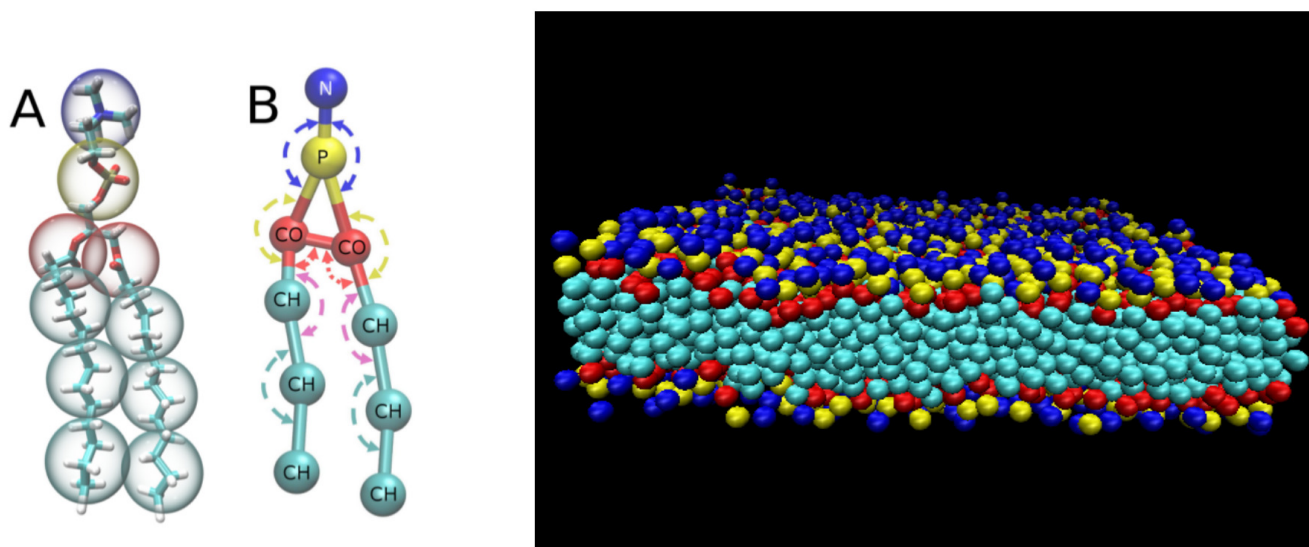


FIG. 12: Simulation snapshot of a single coarse-grained DMPC lipid molecule (left) and of self-assembled DMPC bilayer 15x15 nm containing 762 lipids (right).

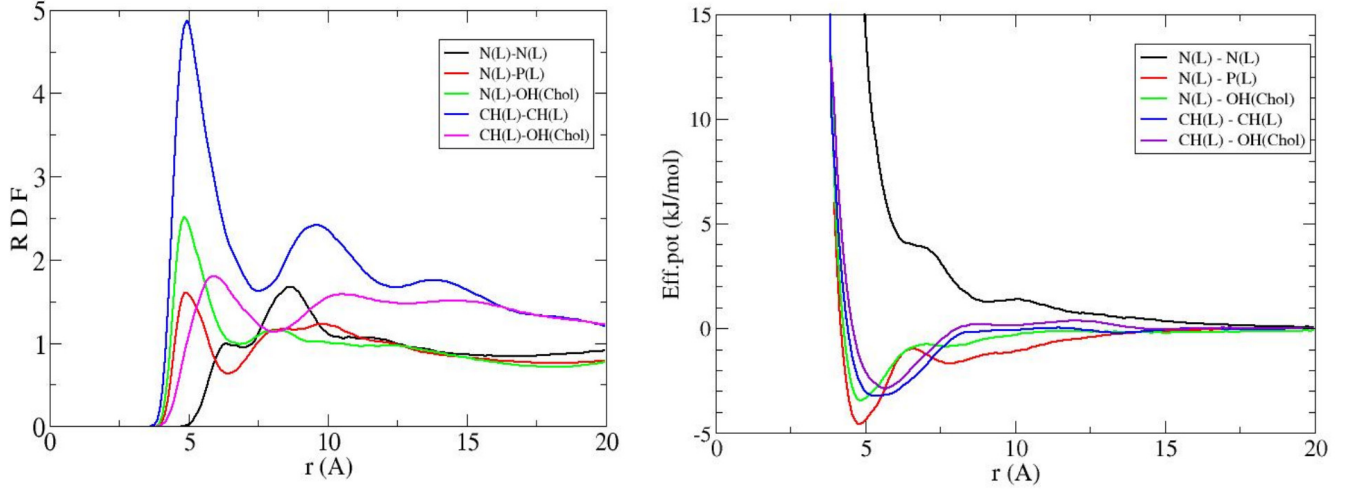


FIG. 13: Extraction via Inverse Monte Carlo of CG site-site pair potentials from atomistic atomistic RDF's. 5 site-site RDFs (left) and corresponding effective potentials involving DMPC CG sites of total 28 for DMPC-CHOL mixture (right) are shown.

	Area per lipid ( $\text{\AA}$ )	Compressibility ( $10^{14}$ N/nm)	Tail order parameters	
			(1)	(2)
atomistic	60	1.9	0.57	0.52
CG	59	2.5	0.56	0.52

TABLE IV: Comparison of the properties of the DMPC bilayer obtained from the full atomistic simulations and the CG model.

carried out a number of simulations of flat lipid bilayers composed of CG lipid models representing other lipids which can be built from the CG sites presented in Fig. 11. These simulations were carried out at zero-tension conditions within the atomistic and coarse-grained models. Table III C shows comparison of some properties not related to the radial distribution functions obtained within atomistic and coarse-grained simulations of a piece of bilayer composed of 128 DMPC lipids. Very good agreement is observed for the average area per lipid (which is one of the most important parameters for a lipid bilayer) and for the tail order parameter, and a reasonably good agreement for the bilayer compressibility. Especially important is the agreement for the order parameter, which shows that orientational fluctuations of the lipid tails are the same in atomistic and coarse-grained models.

Table III C shows average areas per lipid obtained in coarse-grained simulations carried out in conditions of zero tension and experiment for a number of lipids. Except DMPC, other lipids included in this table were not used in the parameterization of the CG potentials. The models for these lipids were built from appropriate sites of DMPC and DOPS lipids shown in Fig. 11, and the CG interaction potentials were taken as determined in IMC computations for DMPC and DOPS lipids (some of them shown in Fig. 13). One can see generally good agreement with experiment, though simulations show a tendency for some underestimation of the lipid area. The bilayer composed of DSPC lipids was found in the gel phase which again is in agreement with experiment (the temperature of gel phase transition for DSPC is  $55^\circ\text{C}$ ). We are not aware of an experimental value of the average lipid area for the gel phase of DSPC, but it is generally accepted that average area per lipid in the gel phase is in the range  $43\text{ \AA} - 48\text{ \AA}$  for phosphatidylcholine lipids. Also, atomistic simulations of DSPC bilayer in a gel phase [54] reports the average lipid area of  $44.5\text{ \AA}^2$  which is in good agreement with the result of our coarse-grained model.

#### IV. NP AND BILAYER SIMULATION

Using the methodologies described in sections II and III we now can construct a model to simulate the interaction of a DPMC lipid bilayer with a small hydrophobic NP and a hydrophobic NP associated with one molecule of HSA. Following is the description of the simulations and the main results.



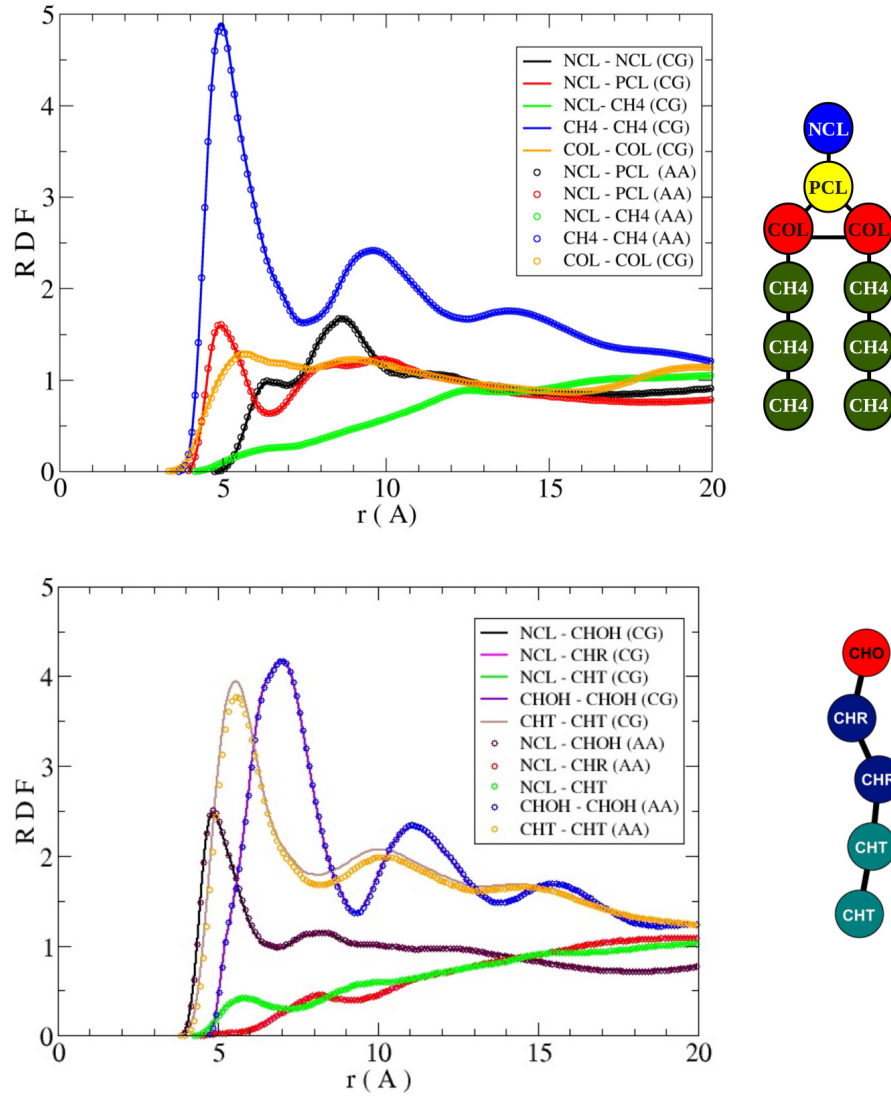


FIG. 14: Radial distribution functions between different sites of DMPC (top graph) and Cholesterol (bottom graph) molecules, see site definitions to the right of the graph. Atomistic (points) and coarse-grained (lines) simulations were carried out for a mixture of 30 DMPC, 30 Cholesterol, and 1800 water molecules (CG simulation were without explicit water but in a box of the same size as atomistic simulations including water).

Lipid	Area per lipid ( $\text{\AA}^2$ )	
	sim	exp
DMPC (14:0/14:0 PC)	59.0	60.5 [55]
SOPC (18:0/18:1n9 PC)	60.4	61.1 [56]
DOPC ( 18:1n9/18:1n9 PC)	62.0	67.4 [57]
DSPC (18:0 /18:0 )	43.5 <sup>1</sup>	44.5 <sup>1,2</sup> [54]

TABLE V: Average areas per lipid. Comparison of simulation results computed in coarse-grained simulations and experiments at T=303K

<sup>1</sup> - bilayer in gel phase <sup>2</sup> - evaluated in atomistic MD simulations

### A. Interaction potentials and parameters of the simulations

The simulated systems are composed of the NP, the lipids that form the bilayer, the amino acids of the HSA, and monovalent ions that are used to resemble physiological conditions. For all interactions we assume two contributions: electrostatic and van der Waals interactions. For the electrostatic contribution, all charged beads interact through a Coulomb potential between given by:

$$U^C(r_{ij}) = \lambda_B k_B T \frac{q_i q_j}{r_{ij}} \quad (9)$$

where  $r_{ij}$  is the distance between the bead  $i$  and the bead  $j$ ,  $\lambda_B$  is the Bjerrum length and  $q_i$  and  $q_j$  are the charges of the beads  $i$  and  $j$  respectively. The calculation of this long-range interaction was implemented through a Ewald summation P3M algorithm [41]. The Bjerrum length is set to 0.71 nm in all cases.

For the van der Waals interactions we use the following model:

- Aminoacid – aminoacid van der Waals interactions: we do not explicitly consider interaction between any pair of aminoacids within the single protein molecule as the protein is not allowed to change conformation from the PDB crystal structure. To improve the computational efficiency instead of simulating the HSA molecule as a rigid body, we connect all residues which are separated less than 10 nm by harmonics bonds with a spring constant of  $100k_B T$ . Fig. 15a shows the resulting CG model for the HSA (build according to PDB ID:1N5U), while Fig. 15b shows the resulting network of bonds (8059 in total).
- NP – aminoacid van der Waals interactions: we used the interaction potential defined in Eq. (2) and the same parameters for the aminoacids and the NP obtained by the parametrization described in Sec. II C.
- Lipid-lipid interactions: the CG models of the lipids in the bilayer and the interactions between the four different types of beads were obtained as described in Sec. III.
- Lipid-NP interactions: we used the same interaction potentials as in the case of NP–aminoacid interaction (Eq. (2)) and assumed that lipid beads interact with the surface according to their hydrophobicity. We classify the lipid beads into one of two groups: head or tail. The head beads are NCL, PCL and COL (see Fig. 11), and they are considered to be hydrophilic with a value of  $\epsilon_i = 0.1$ . The tail beads (labeled CH4 in Fig. 11) are hydrophobic and a value of  $\epsilon_i = 0.75$  is used for these group. The van der Waals radius of all the lipids beads is set to  $\sigma_i = 0.6$  nm.
- Lipid-aminoacid interactions: for these interactions we use the same approach as for the NP-residues and NP-lipid. The potential interaction is also based on the hydrophobicity of the beads given by the following modified 12-6 Lennard-Jones potential:

$$U_{l,i}(r) = \begin{cases} 4\epsilon_{la} \left[ \left( \frac{\sigma_{l,i}}{r} \right)^{12} - \left( \frac{\sigma_{l,i}}{r} \right)^6 \right] + \epsilon_{la}(1 - \epsilon_{l,i}) & r < r_c \\ 4\epsilon_{la}\epsilon_{l,i} \left[ \left( \frac{\sigma_{l,i}}{r} \right)^{12} - \left( \frac{\sigma_{l,i}}{r} \right)^6 \right] & r_c \leq r \leq r_{\text{cut}} \\ 0 & r > r_{\text{cut}} \end{cases} \quad (10)$$

where  $r$  is the distance from the lipid bead  $l$  to the residue  $i$ ,  $\epsilon_{la}$  is a free parameter that scales the interaction energy,  $\epsilon_{l,i}$  is the combined hydrophobicity index of lipid  $l$  and the residue  $i$  and is given by  $\epsilon_{l,i} = \sqrt{\epsilon_l \epsilon_i}$ ,  $\sigma_{l,i}$  is the average van der Waals radius of residue  $i$  and the lipid  $l$ ,  $\sigma_{l,i} = (\sigma_l + \sigma_i)/2$ ,  $r_c = 2^{1/6} \sigma_{l,i}$  and  $r_{\text{cut}}$  is the cut-off for the van der Waals interaction. As in this work we only study the applicability of the proposed methodology, we do not systematically parameterize the value of  $\epsilon_{la}$ , instead we set this scaling parameter to  $0.5k_B T$  for all simulations. This values gives interactions between the lipids and the residues in the same order of magnitude as the ones reported in [58].

- Ion-ion and ion-molecule interactions: in addition to the Coulomb forces, we include excluded volume interactions by means of a WCA potential. The van der Waals radii of the ions are set to 0.2 nm.

For all simulations a NP of radius 2 nm is used with a surface charge of  $-0.02$  C/m<sup>2</sup>. We use *NVT* ensemble with the box size was  $15 \times 15 \times 20$  nm and we assume physiological conditions with monovalent salt concentration of 0.1 M (270 negative and positive ions are placed in the simulation box). To keep charge neutrality, further 16 positive ions are added. The bilayer is composed of a total of 762 lipids (381 lipids in each layer). A Langevin thermostat with a friction coefficient of  $\gamma = 0.05$  is used and the units of mass, energy and charge are the same as described in Sec. II C. The time unit ( $\tau$ ) is obtained by performing a simulation of the bilayer with ions (no NP or proteins are added) and measuring the lateral diffusion constant. We obtained a value of  $8 \times 10^{-5}$  nm<sup>2</sup>/ $\tau$ , which compared with the experimental value of  $5 \mu\text{m}^2/\text{s}$  gives  $\tau = 16$  ps.

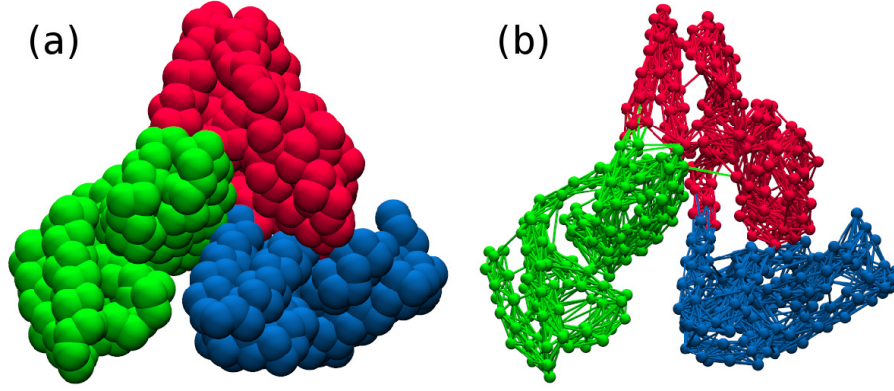


FIG. 15: (a) CG model of HSA protein used in this study. Each residue is represented by a single bead located at the position of the  $\alpha$ -carbon. Each color represents one of the three domains of the HSA molecule. (b) All beads separated by less than 10 nm are connected by stiff harmonic bonds.

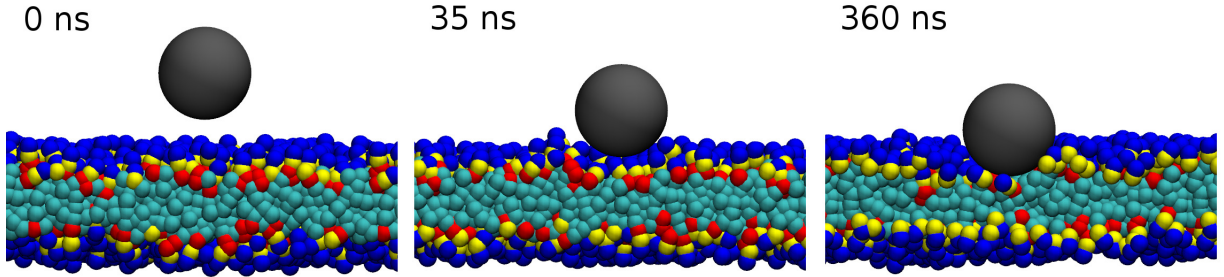


FIG. 16: Time sequence of the snapshots of the interaction of a DMPC lipid bilayer with a negative charged hydrophobic NP. The radius of the NP is 2 nm. The ions are not shown.

### B. Simulation of a nanoparticle in contact with a lipid bilayer

To study the interaction of a bare NP with the lipid bilayer we initially position the NP close to the bilayer surface and follow the time dynamics of the system. Three snapshots are shown Fig. 16. From the initial state the NP adsorbs quickly to the surface of the bilayer and then penetrates to around 1 nm inside the membrane and stays strongly attached until the end of the simulation (see snapshots at 35 and 360 ns). To explore the adsorption process in more detail, the distance of the surface of the NP to the center of the bilayer was recorded and the results are shown in Fig. 17. The NP reaches the bilayer in a few nanoseconds and then attaches to the surface for around 50 ns (the dashed line in Fig. 17 marks the average position of the surface of the lipid bilayer, as defined by the position of the maximum of density of the lipid headgroups). After that, the NP starts penetrating the membrane (in a few nanoseconds). Then a slow internalization is observed until the NP reaches its final position at approximately 300 ns. The internalization of the NP is mediated by the attractions between both type of lipids and the NP. The penetration then stops (or becomes much slower) because any further displacement requires a substantial change in the membrane structure. To study long-time dynamics of the system, NP lipid wrapping and uptake one needs a bigger bilayer or/and NPT ensemble [59]. Despite of this limitation, the results obtained with our methodology agree with a recent report [60] for the absorption of a hydrophobic NP with a membrane composed of lipids and specialized receptors.

### C. Simulation of a nanoparticle-protein complex complex in contact with a lipid bilayer

It is well accepted that in more realistic situations, the NP before it reaches the cell membrane gets coated by proteins and that this NP-protein complex is responsible for the final fate of the NP [61]. Considering this, we now simulate the interaction of a NP-protein complex, where protein is represented by a single HSA molecule. As shown

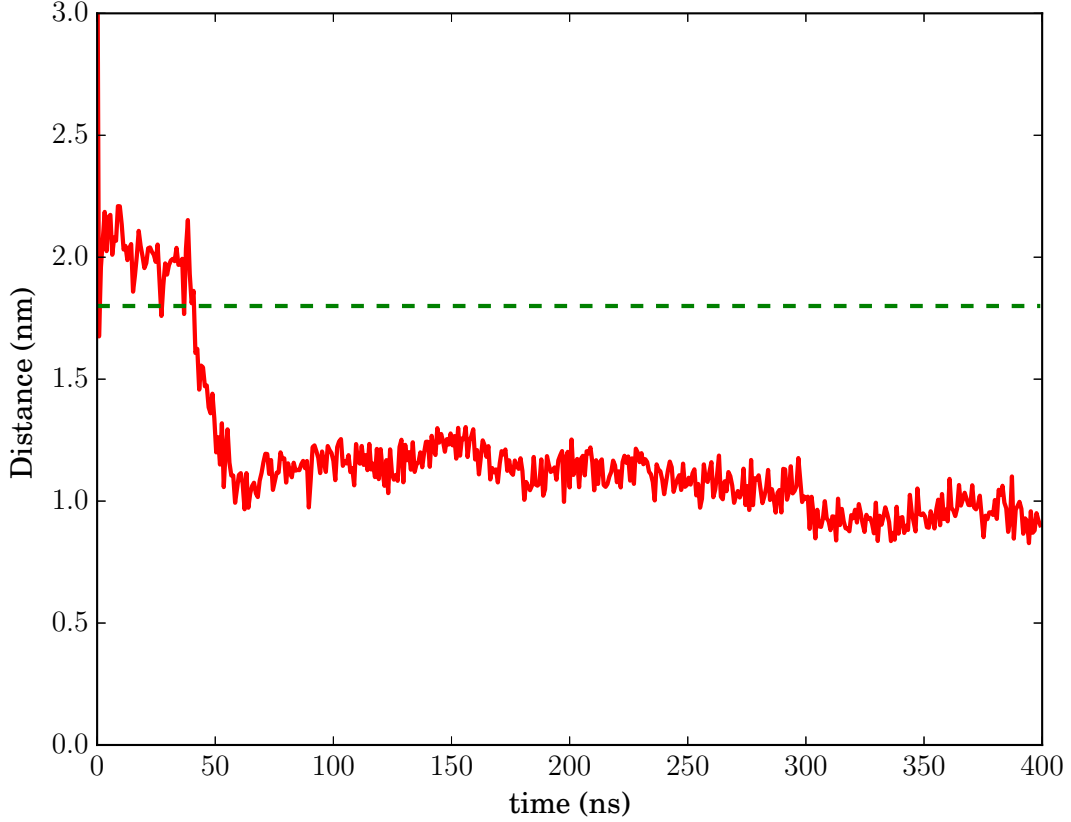


FIG. 17: Distance of the surface of the bare NP to the center of the bilayer as a function of time. The dashed line shows the average position of the bilayer surface.

in Sec. II, not all orientations in which protein adsorbs onto a surface are equally probable and for our simulation of the interaction of a hydrophobic NP with a DPMC lipid membrane we first calculate the adsorption energy map of HSA onto a 2 nm of radius hydrophobic NP. The adsorption map is shown in Fig. 18a. We can see that, as in the cases discussed above, the energy landscape contains more than one minimum. We found the average adsorption energy of  $-1.7 k_B T$  and as the initial orientation for the simulation we selected the orientation  $\theta, \phi = 145^\circ, 110^\circ$ , which corresponds to adsorption energy of  $-2.7 k_B T$  and the corresponding complex NP-HSA is shown in Fig. 18b.

Figure 19 shows a sequence of snapshots from simulation of the NP-HSA complex with the lipid bilayer. In the initial state, the protein is facing the bilayer. In the simulation, the HSA at first moves in front of the membrane and prevents a direct contact between the NP and the lipids. Then, the NP-HSA complex rotates so that the NP faces the bilayer and starts penetrating the membrane. Fig. 20 shows the distance of the NP surface to the center of the membrane. The rotation is reflected in the sudden change of the position of COM of the HSA. After this quick rearrangement the NP starts the penetration while the protein stays attached to the NP for the whole simulation but moves around the surface of the NP as can be seen in the snapshot for the times 140 and 300 ns in Fig. 19. This movement of the HSA molecule can also be observed from the curve of the COM of the HSA curves as a function of time (Fig. 20).

Comparing the two simulations we see that the presence of the HSA dramatically changes the interaction of the NP with the membrane. We can envision that a NP, fully coated NP with HSA, will not be able to penetrate into the membrane, so the coating is changing completely its biological reactivity.

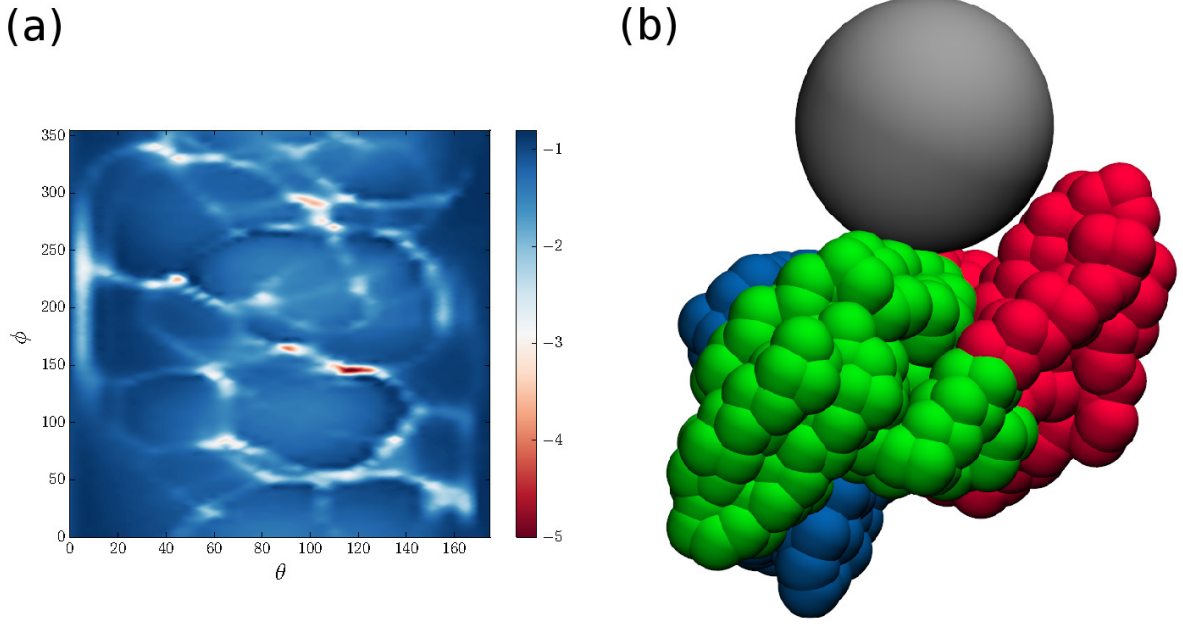


FIG. 18: (a) Initial state of the NP-HSA complex. (b) Surface map of the adsorption orientations of HSA onto a 2 nm negatively charged hydrophobic NP.

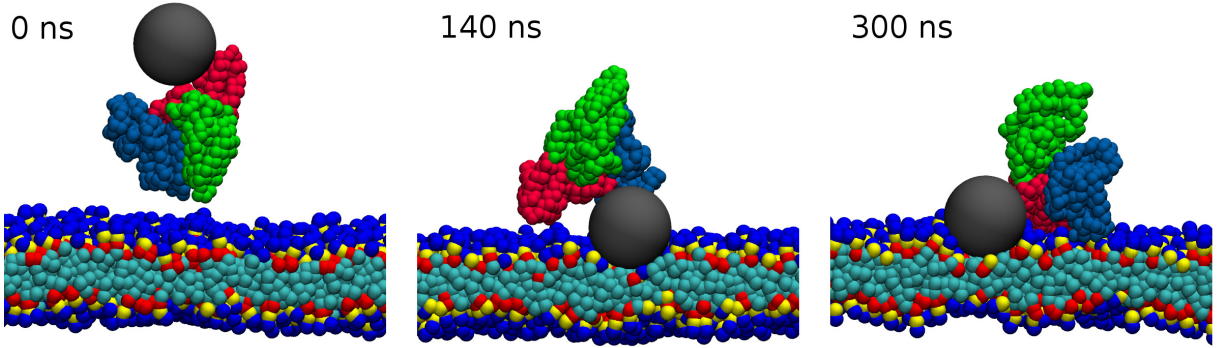


FIG. 19: Time sequence of the snapshots of the interaction of a DMPC lipid bilayer with a negatively charged hydrophobic NP complex with one HSA molecule. The radius of the NP is 2 nm. The ions are not shown.

## V. CONCLUSIONS

In this work, we presented a multiscale methodology for modelling interactions at bionano interface, which is central for understanding uptake and toxicity of nanomaterials. We used systematic coarse-graining techniques to reduced the complexity of the problem by removing some degrees of freedom and focussing on the properties of interest. Since the CG models consists of about ten times less interaction centres than the atomistic model, and the solvent (water) is not modeled explicitly, simulations of the CG model take 2-3 order less CPU time compared with atomistic simulations for equal system size, or, alternatively, CG model can be used for simulations of whole proteins, small NPs and sufficiently large cell membrane fragments at the scale of tens of nanometers. We have parameterised and validated the model against experiments.

The technique for coarse-graining NP-protein interaction, which we presented in Section II can be used to calculate the binding energies for arbitrary plasma, cytosolic or membrane proteins, rank them by binding affinity to the NP and predict the content of NP protein corona. Our calculations show that the NP surface charge has a small effect on the adsorption energies in comparison to van der Waals interactions between the residues and the surface. We also find that the charge of the NP does not influence much the orientation, in which the protein prefer to adsorb. On

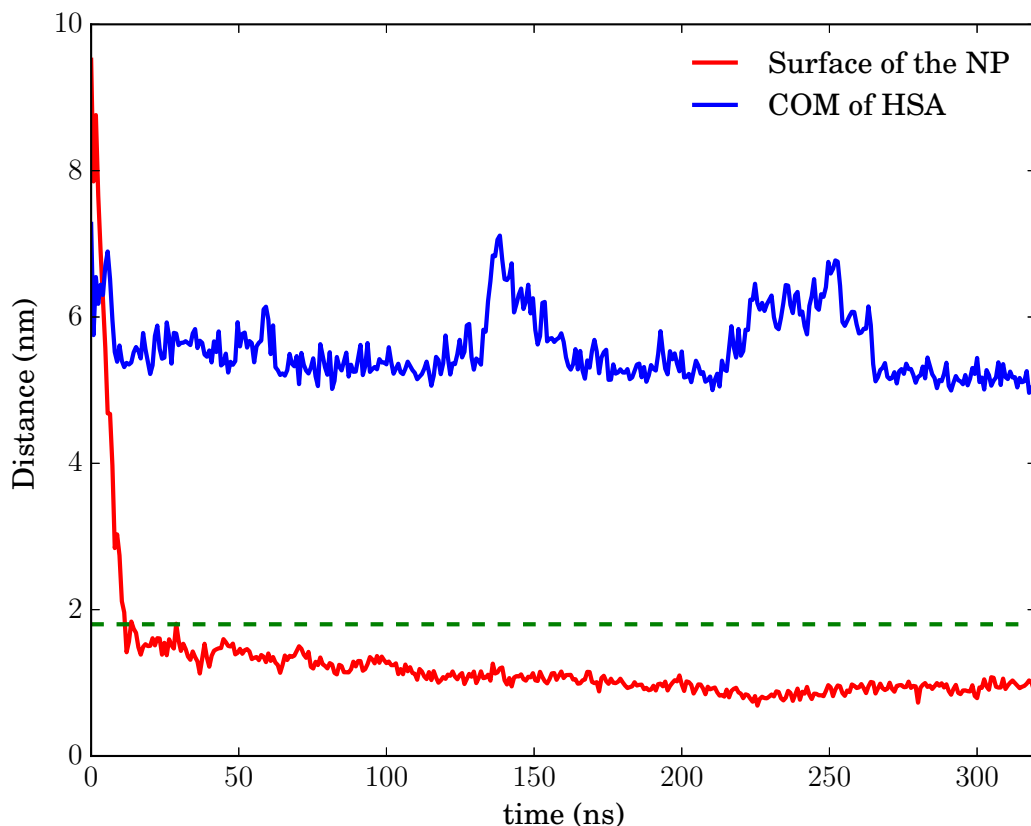


FIG. 20: Distance of the surface of the complex of NP with one HSA to the center of the bilayer and the COM of the HSA as a function of time. The dashed line shows the average position of the bilayer surface.

the other hand, we have shown the size of the NP has a big effect on the adsorption energy maps, as the curvature of the NP determine the sections of the protein that can interact with the surface. Based on our simulations results, we can predict bigger proteins adsorb stronger on the inorganic surfaces, even for small NPs, in agreement with the Vroman effect. We have also demonstrated that a rigid protein model is justified at least for small globular proteins. In Section III, we have parameterised a CG lipid and cholesterol model, which reproduce the key bilayer properties of atomistic model of the same system. Finally, in Section IV, we have shown how the CG lipid and NP-protein models can be combined to model NP-cell membrane interactions and NP attachment and uptake.

- 
- [1] S. Sharifi, S. Behzadi, S. Laurent, M.L. Forrest, P. Stroeve, and M. Mahmoudi. Toxicity of nanomaterials. *Chem. Soc. Rev.*, 41:2323, 2012.
  - [2] P. J. A. Borm and et al. The potential risks of nanomaterials. *Part. Fibre Toxicol.*, 3:11, 2006.
  - [3] H.J. Johnston, G.R. Hutchison, F.M. Christensen, and et al. Identification of the mechanisms that drive the toxicity of tio2 particulates: the contribution of physicochemical characteristics. *Part. Fibre Toxicol.*, 6:33, 2009.
  - [4] A. E. Nel, L. Maedler, D. Velegol, and et al. Understanding biophysicochemical interactions at the nano-bio interface. *Nat. Mater.*, 8:543, 2009.
  - [5] A. Verma, O. Uzun, Y. Hu, and et al. Surface-structure-regulated cell-membrane penetration by monolayer-protected nanoparticles. *Nat. Mater.*, 7:588, 2008.
  - [6] T. Schlick, R. Colleparado-Guevara, L.A. Halvorsen, and et al. Biomolecular modeling and simulation: a field coming of age. *Quarterly Rev. Biophys.*, 44:191, 2011.
  - [7] L.G. Valerio Jr. In silico toxicology for the pharmaceutical sciences. *Toxicol. Appl. Pharmacol.*, 241:356, 2009.
  - [8] F. Nigsch, N.J. Macaluso, J. B. Mitchell, and D. Zmuidinavicius. Computational toxicology: an overview of the sources of data and of modelling methods. *Expert Opin. Drug Metab. Toxicol.*, 5:1, 2009.



- [9] J.C. Dearden. In silico prediction of drug toxicity. *J. Comput.-Aided Mol. Des.*, 17:119, 2003.
- [10] A. P. Lyubartsev and A. L. Rabinovich. Recent development in computer simulations of lipid bilayers. *Soft Matter*, 7:25, 2011.
- [11] J. Wong-Ekkabut, S. Baoukina, W. Triampo, I.-M. Tang, and D. P. Tieleman. Computer simulation study of fullerene translocation through lipid membranes. *Nat. Nanotechnol.*, 3:363, 2008.
- [12] W.C. Hou, B. Y. Moghadam, P. Westerhoff, and J. D. Posner. Distribution of fullerene nanomaterials between water and model biological membranes. *Langmuir*, 27:11899, 2011.
- [13] K. Yang and Y. Q. Ma. Computer simulation of the translocation of nanoparticles with different shapes across a lipid bilayer. *Nat. Nanotechnol.*, 5:579, 2010.
- [14] L. Monticelli, E. Salonen, P.C. Ke, and I. Vattulainen. Effects of carbon nanoparticles on lipid membranes: a molecular simulation perspective. *Soft Matter*, 5:4433, 2009.
- [15] S. Izvekov and G.A. Voth. Multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, 109:2469, 2005.
- [16] G. S. Ayton, W. G. Noid, and G. A. Voth. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.*, 17:192, 2007.
- [17] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Phys. Rev. E*, 52:3730, 1995.
- [18] A. P. Lyubartsev, A. Mirzoev, L.-J. Chen, and A. Laaksonen. Systematic coarse-graining of molecular models by the newton inversion method. *Faraday Discuss.*, 144:43, 2010.
- [19] A. P. Lyubartsev and A. Laaksonen. Effective potentials for ion-dna interactions. *J. Chem. Phys.*, 111:11207, 1999.
- [20] V. Tozzini. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.*, 15(2):144–150, 2005.
- [21] T. Bereau and M. Deserno. Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.*, 130(23):235106, 2009.
- [22] S. Takada. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.*, 22(2):130–137, 2012.
- [23] S. Wei and T. Knotts. A coarse grain model for protein-surface interactions. *J. Chem. Phys.*, 139(9):095102, 2013.
- [24] V. Lobaskin, A. P. Lyubartsev, and P. Linse. Effective macroion-macroion potentials in asymmetric electrolytes. *Phys. Rev. E*, 63:020401, 2001.
- [25] M. Brunner, C. Bechinger, W. Strepp, V. Lobaskin, and H. H. von Gruenberg. Density-dependent pair interactions in 2d colloidal dispersions. *Europhys. Lett.*, 58:926, 2002.
- [26] I. Lynch, A. Salvati, and K. A. Dawson. Protein-nanoparticle interactions. what does the cell see? *Nat. Nanotechnol.*, 4:546, 2009.
- [27] I. Lynch, K. A. Dawson, and S. Linse. Detecting cryptic epitopes created by nanoparticles. *Sci. STKE2006*, page pe14, 2006.
- [28] T. Cedervall and et al. Understanding the nanoparticle protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. *Proc. Natl. Acad. Sci. U.S.A.*, 104:2050, 2007.
- [29] S. Lindman and et al. Systematic investigation of the thermodynamics of hsa adsorption to n-iso-propylacrylamide/n-tert-butylacrylamide copolymer nanoparticles. effects of particle size and hydrophobicity. *Nanoletters*, 7:914, 2007.
- [30] L. T. Allen and et al. Surface-induced changes in protein adsorption and implications for cellular phenotypic responses to surface interaction. *Biomaterials*, 27:3096, 2006.
- [31] C. E. Radke and J. M. Prausnitz. Thermodynamics of multisolute adsorption from dilute liquid solutions. *AIChE J.*, 18:761, 1972.
- [32] A. Lesniak, A. Campbell, M. P. Monopoli, I. Lynch, A. Salvati, and K. A. Dawson. Serum heat inactivation affects protein corona composition and nanoparticle uptake. *Biomaterials*, 31:9511, 2010.
- [33] W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.*, 139(9):090901, 2013.
- [34] H. Lopez and V. Lobaskin. Coarse-grained model of adsorption of blood plasma proteins onto nanoparticles. *arxiv:1508.01450*, 2015.
- [35] S. Miyazawa and R.L. Jernigan. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, 256(3):623–644, 1996.
- [36] Y. Kim, C. Tang, G. Clore, and G. Hummer. Replica exchange simulations of transient encounter complexes in protein-protein association. *Proc. Natl. Acad. Sci. USA*, 105(35):12855–12860, 2008.
- [37] Y. Kim and G. Hummer. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.*, 375(5):1416–1433, 2008.
- [38] M. Agashe, V. Raut, S. Stuart, and R. Latour. Molecular simulation to characterize the adsorption behavior of a fibrinogen  $\gamma$ -chain fragment. *Langmuir*, 21(3):1103–1117, 2005.
- [39] Y. Sun, W. Welsh, and R. Latour. Prediction of the orientations of adsorbed protein using an empirical energy function with implicit solvation. *Langmuir*, 21(12):5616–5626, 2005.
- [40] D. Kokh, S. Corni, P. Winn, M. Hoefling, K. Gottschalk, and R. Wade. Prometcs: An atomistic force field for modeling protein-metal surface interactions in a continuum aqueous solvent. *J. Chem. Theory Comput.*, 6(5):1753–1768, 2010.
- [41] H. Limbach, A. Arnold, B. Mann, and C. Holm. ESPResSo - an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.*, 174(9):704–727, 2006.
- [42] W. Chen, H. Huang, C. Lin, F. Lin, and Y. Chan. Effect of temperature on hydrophobic interaction between proteins and hydrophobic adsorbents: Studies by isothermal titration calorimetry and the van’t hoff equation. *Langmuir*, 19(22):9395–9403, 2003.
- [43] S. Lacerda, Jung J. Park, C. Meuse, D. Pristinski, M. Becker, A. Karim, and J. Douglas. Interaction of gold nanoparticles with common human blood proteins. *ACS Nano*, 4(1):365–379, 2010.

- [44] P. Vilaseca, K. Dawson, and G. Franzese. Understanding and modulating the competitive surface-adsorption of proteins through coarse-grained molecular dynamics simulations. *Soft Matter*, 9:6978–6985, 2013.
- [45] S. Vijay-Kumar, C. Bugg, and W. Cook. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.*, 194:531–544, 1987.
- [46] K. Momma and F. Izumi. *VESTA3* for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of Applied Crystallography*, 44:1272–1276, 2011.
- [47] E. G. Brandt and A. Lyubartsev. Systematic optimization of a force field for classical simulations of interfaces between  $\text{tio}_2$  surfaces and water. In preparation.
- [48] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [49] W. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.
- [50] S. Nose. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.
- [51] J. P. M. Jämbek and A. P. Lyubartsev. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J. Phys. Chem. B*, 116(10):3164–3179, 2012.
- [52] J. P. M. Jämbek and A. P. Lyubartsev. Another piece of the membrane puzzle: Extending slipids further. *J. Chem. Theory Comput.*, 9(1):774–784, 2013.
- [53] A. Mirzoev and A. P. Lyubartsev. Magic: Software package for multiscale modeling. *J. Chem. Theory Comput.*, 9(3):1512–1520, 2013.
- [54] S.-S. Qin, Z. W. Yu, and Y.-X. Yu. Structural characterization on the gel to liquid-crystal phase transition of fully hydrated DSPC and DSPE bilayers. *J. Phys. Chem. B*, 113:8114 – 8123, 2009.
- [55] N. Kučerka, Y. Liu, N. Chu, H. I. Petrache, S. Tristram-Nagle, and J. F. Nagle. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.*, 88:2626 – 2637, 2005.
- [56] B. W. Köenig, H. H. Strey, and K. Gawrisch. Membrane lateral compressibility determined by NMR and X-ray diffraction: effect of acyc chain polyunsaturation. *Biophys. J.*, 73(4):1954 – 1966, 1997.
- [57] N. Kučerka, J. F. Nagle, J. N. Sachs, S. E. Feller, J. Pencer, A. Jackson, and J. Katsaras. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-Ray scattering data. *Biophys. J.*, 95(5):2356 – 2367, 2008.
- [58] T. Bereau, Z.-J. Wang, and M. Deserno. More than the sum of its parts: Coarse-grained peptide-lipid interactions from a simple cross-parametrization. *J. Chem. Phys.*, 140(11):115101, 2014.
- [59] J. Lin, H. Zhang, Z. Chen, and Y. Zheng. Penetration of lipid membranes by gold nanoparticles: Insights into cellular uptake, cytotoxicity, and their relationship. *ACS Nano*, 4(9):5421–5429, 2010.
- [60] Hong-Ming D. and Yu-Qiang M. Computer simulation of the role of protein corona in cellular delivery of nanoparticles. *Biomaterials*, 35(30):8703 – 8710, 2014.
- [61] M. Monopoli, C. Aberg, A. Salvati, and Kenneth A. Dawson. Biomolecular coronas provide the biological identity of nanosized materials. *Nat. Nanotechnol.*, 7(12):779–786, 2012.